

2019

A comparison of prescriptive usage problems in formal and informal written English

Tyler Jordan Smith
Iowa State University

Follow this and additional works at: <https://lib.dr.iastate.edu/etd>



Part of the [Communication Commons](#), [Linguistics Commons](#), and the [Rhetoric Commons](#)

Recommended Citation

Smith, Tyler Jordan, "A comparison of prescriptive usage problems in formal and informal written English" (2019). *Graduate Theses and Dissertations*. 17566.
<https://lib.dr.iastate.edu/etd/17566>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

A comparison of prescriptive usage problems in formal and informal written English

by

Tyler Jordan Smith

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Co-majors: Rhetoric and Professional Communication; Applied Linguistics and Technology

Program of Study Committee:

Bethany Gray, Co-major Professor

Jo Mackiewicz, Co-major Professor

Tracy Lucht

David R. Russell

Sowmya Vajjala

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2019

Copyright © Tyler Jordan Smith, 2019. All rights reserved.

DEDICATION

To Melissa, Day, Neve, and Kit.

For all your support, understanding, and love.

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	viii
ACKNOWLEDGMENTS	x
ABSTRACT	xi
CHAPTER 1. INTRODUCTION	1
1.1 Register, Style, and Formality: A Theoretical Framework	7
1.1.1 Register	8
1.1.2 Style	11
1.1.3 Formality	13
1.2 Contribution of the Study	18
1.3 Overview of the Study	19
CHAPTER 2. LITERATURE REVIEW	21
2.1 Prescriptivism as an Area of Linguistic Interest	21
2.2 Prescriptivism as a Social Phenomenon	24
2.3 Origins of Prescriptivism	27
2.4 Prescriptivism Today	28
2.5 Usage Problems and Usage Guides	30
2.6 The Interaction between Prescriptivism and Usage	33
CHAPTER 3. REGISTER ANALYSIS	37
3.1 Justifying the Choice of Personal Blogs and News	37
3.2 Comparing the Situational Formality of Personal Blogs and News	39
3.3 Changes in the Production of News Writing	41
3.4 Research Questions	45
CHAPTER 4. RESEARCH DESIGN	47
4.1 Determining the Level of Prescriptivism in Usage Guides (RQ1)	47
4.1.1 Selecting the Usage Problems to Include in the Study	48

4.1.2 Selecting the Usage Guides to Include in the Study.....	50
4.1.3 Compiling the Relevant Entries from the Usage Guides into a Rating Sheet.....	51
4.1.4 Developing a Scale to Calculate a Prescriptivism Index	56
4.1.5 Using the Rating Scale to Calculate a Prescriptivism Index for Each Usage Problem in Each Usage Guide.	58
4.2 Conducting the Corpus Analysis (RQ2).....	61
4.2.1 COCA–N.....	61
4.2.2 Corpus of Personal Blog Posts (CPBP)	61
4.2.2.1 Collecting URLs	62
4.2.2.2 Determining author location and status as personal blog.....	66
4.2.2.3 Scraping, cleaning, and tagging the texts.....	73
4.2.2.4 Determining the purpose and topic of texts.....	75
4.2.3 Extracting Data from Both Corpora	79
4.2.3.1 Determining whether to use specific or general search terms.....	80
4.2.3.2 Determining the number of instances to sample.....	81
4.2.3.3 Sampling the data and creating the dataset	82
4.2.3.4 Nature of the data.....	82
4.3 Administering the Survey to Bloggers and Journalists (RQ3).....	85
CHAPTER 5. RESULTS	89
5.1 LAY/LIE.....	89
5.1.1 Review of Advice from Usage Guides	90
5.1.2 Research Questions and Method of Extracting Data from the Corpora	92
5.1.3 Results of Corpus Analysis	95
5.1.4 Summary of Survey Results.....	98
5.2 WHO/WHOM.....	100
5.2.1 Level of Prescriptivism in Usage Guides.....	100
5.2.2 Research Questions and Method of Extracting Data from the Corpora	102
5.2.3 Results of Corpus Analysis	104
5.2.4 Summary of Survey Results.....	107
5.3 DIFFERENT TO/THAN/FROM.....	110
5.3.1 Level of Prescriptivism in Usage Guides.....	110
5.3.2 Research Questions and Method of Extracting Data from the Corpora	112
5.3.3 Results of Corpus Analysis	113
5.3.4 Summary of Survey Results.....	113

5.4 SPLIT INFINITIVE	116
5.4.1 Review of Advice from Usage Guides	116
5.4.2 Research Questions and Method of Extracting Data from the Corpora	118
5.4.3 Results of Corpus Analysis	119
5.4.4 Summary of Survey Results	120
5.5 I FOR ME	122
5.5.1 Level of Prescriptivism in Usage Guides	123
5.5.2 Research Questions and Method of Extracting Data from the Corpora	124
5.5.3 Results of Corpus Analysis	125
5.5.4 Summary of Survey Results	126
5.6 SINGULAR THEY	128
5.6.1 Level of Prescriptivism in Usage Guides	129
5.6.2 Research Questions and Method of Extracting Data from the Corpora	131
5.6.3 Results of Corpus Analysis	135
5.6.4 Summary of Survey Results	135
5.7 LESS/FEWER	138
5.7.1 Level of Prescriptivism in Usage Guides	138
5.7.2 Research Questions and Method of Extracting Data from the Corpora	139
5.7.3 Results of Corpus Analysis	140
5.7.4 Summary of Survey Results	141
5.8 NONE IN PLURAL CONTEXT	143
5.8.1 Level of Prescriptivism in Usage Guides	144
5.8.2 Research Questions and Method of Extracting Data from the Corpora	144
5.8.3 Results of Corpus Analysis	147
5.8.4 Summary of Survey Results	148
5.9 Discussion of Results	150
5.9.1 Comparison of Patterns of Usage	150
5.9.2 Attitudes of Usage-Guide Authors, Bloggers, and News Writers Toward the Usage Problems	156
5.9.3 Alignment of Usage Patterns with Attitude Profiles	159
5.9.4 Profiles of Usage Guides	161
CHAPTER 6. CONCLUSIONS AND IMPLICATIONS FOR TECHNICAL EDITING ...	163
6.1 Review of Research Questions	163
6.1.1 RQ1	163

6.1.2 RQ2.....	164
6.1.3 RQ3.....	165
6.2 Implications for Technical Editing.....	168
6.2.1 Favoring a Rhetorical Approach to Technical Editing.....	168
6.2.2 Using Empirical Data to Inform a Rhetorical Approach to Technical Editing....	171
6.2.3 Describing the Benefits of a Rhetorical Approach to Technical Editing.....	174
REFERENCES	182
APPENDIX A: UCREL CLAWS7 TAGSET	192
APPENDIX B: SURVEY INSTRUMENT.....	196
APPENDIX C: PRESCRIPTIVISM RATINGS FOR EACH USAGE PROBLEM	209
APPENDIX D: WALSH AND WALSH’S (1989) LIST OF GRAMMATICAL PATTERNS FOR <i>WHO/WHOM</i>	210
APPENDIX E: REGULAR EXPRESSIONS USED TO EXTRACT DATA FOR SINGULAR THEY	211

LIST OF TABLES

Table 4.1 Usage problems included in the current study.....	49
Table 4.2 Comparison of usage guides in the HUGE database and those included in this study	52
Table 4.3 Excerpts from the usage-guide entries that exemplify each assigned index	60
Table 4.4 Trigrams used in online searches.....	65
Table 4.5 Summary of the steps taken to collect the blog URLs used in this study	73
Table 4.6 Composition of corpora.....	79
Table 4.7 Variants analyzed for each usage problem and the number of variants analyzed in both corpora.....	83
Table 4.8 Framework developed to classify prescriptive rules.....	85
Table 4.9 Age ranges and genders of survey participants	87
Table 4.10 Geographic locations of survey participants	88
Table 5.1 Inflectional patters of <i>lay</i> and <i>lie</i>	89
Table 5.2 Search terms used to extract data from both corpora.....	92
Table 5.3 Search terms used to extract data from both corpora.....	103
Table 5.4 Examples of <i>who</i> and <i>whom</i> from each of the grammatical positions in which they were observed	103
Table 5.5 Percent of prescriptively correct and incorrect variants by grammatical category	106
Table 5.6 Search term used to find instances of <i>different than</i>	112
Table 5.7 Search terms used to extract data from both corpora.....	119
Table 5.8 Search terms used to extract data from both corpora.....	125
Table 5.9 Third-person pronouns to consider when studying variation in singular <i>they</i>	133
Table 5.10 Search terms used to extract data from both corpora.....	140
Table 5.11 Search terms used to extract data from both corproa.....	145
Table 5.12 Classification of usage problems according to formal/informal distinctions.....	154
Table 5.13 Perceived level of formality of blog writing and news writing.....	155
Table 5.14 Average prescriptivism index for each usage guide analyzed for this study	162

LIST OF FIGURES

Figure 4.1 Excerpt of the rating instrument used to rate the level of prescriptivism in usage guides	56
Figure 4.2 Scale created to rate the level of prescriptivism for each usage problem in each usage guide	58
Figure 4.3 Map of American dialect regions	68
Figure 4.4 Homepage of a personal blog included in this study.....	72
Figure 4.5 Homepage of an organizational blog excluded from this study.....	72
Figure 4.6 Survey instrument used by Mechanical Turk workers to determine the purpose, topic, and amount of quoted material in a sample of blog posts.....	77
Figure 5.1 Portion of the coding instrument used to analyze instances of <i>lay</i> , <i>lie</i> , and their derivatives.....	93
Figure 5.2 Proportion of correctly and incorrectly used instances of <i>lay</i> , <i>lie</i> and their derivatives in blog and news writing.....	96
Figure 5.3 Proportion of errors in each verb group.....	97
Figure 5.4 Proportion of bloggers and journalists who found the misuse of LAY/LIE unacceptable, acceptable only in informal contexts, or acceptable in formal contexts.....	99
Figure 5.5 Proportion of correctly and incorrectly used instances of <i>who</i> and <i>whom</i> in blog writing and news writing	105
Figure 5.6 Raw frequencies of grammatical constructions in which <i>who</i> and <i>whom</i> were observed	107
Figure 5.7 Proportion of bloggers and journalists who found the misuse of WHO/WHOM unacceptable, acceptable only in informal contexts, or acceptable in formal contexts...	109
Figure 5.8 Proportion of instances of <i>different than</i> followed by a clause or a phrase in blog writing and news writing	114
Figure 5.9 Proportion of bloggers and journalists who found the misuse of DIFFERENT TO/THAN/FROM unacceptable, acceptable only in informal contexts, or acceptable in formal contexts.....	115
Figure 5.10 Proportion of split and unsplit infinitives in blog and news writing.....	120
Figure 5.11 Proportion of bloggers and journalists who found the SPLIT INFINITIVE unacceptable, acceptable only in informal contexts, or acceptable in formal contexts...	122
Figure 5.12 Proportion of nominative and accusative conjoined noun phrases that were used correctly and incorrectly in blog and news writing	127
Figure 5.13 Proportion of bloggers and journalists who found the misuse of I FOR ME unacceptable, acceptable only in informal contexts, or acceptable in formal contexts...	128
Figure 5.14 Proportion of singular and plural pronouns with an indefinite pronoun as the antecedent in blog writing and news writing	136

Figure 5.15 Proportion of bloggers and journalists who found <i>they</i> used as a singular pronoun unacceptable, acceptable only in informal contexts, or acceptable in formal contexts	137
Figure 5.16 Proportion of <i>less</i> and <i>fewer</i> used to modify plural countable nouns in blog writing and news writing	141
Figure 5.17 Proportion of bloggers and journalists who found the misuse of LESS/FEWER unacceptable, acceptable only in informal contexts, or acceptable in formal contexts... 142	
Figure 5.18 Proportion of <i>none</i> used in singular or plural contexts in blog and news writing.....	148
Figure 5.19 Proportion of bloggers and journalists who found the <i>none</i> used with a plural verb unacceptable, acceptable only in informal contexts, or acceptable in formal contexts.....	149
Figure 5.20 Comparison of correct and incorrect instances of eight usage problems in blog writing and news writing	152
Figure 5.21 Box and whisker plot showing the descriptive statistics for the prescriptivism index of each usage problem.....	156
Figure 5.22 Comparison of attitude profiles.....	158

ACKNOWLEDGMENTS

The project presented in this dissertation could never have been a reality if not for the help, support, and encouragement of many people. I would first like to thank Jo Mackiewicz and Bethany Gray, my co-major professors, for their mentorship, feedback, encouragement, and everything they taught me; David Russell for his unwavering support and encouragement; Tracy Lucht for her invaluable insights into news writing, a perspective I now realize that is much needed in corpus research; and Sowmya Vajalla for giving me a foundation on which to build my programming skills and for her patient help with the programming roadblocks I encountered.

I would also like to thank those who helped me with the large amount of data collection and coding necessary for this project. Many family members and friends helped me gather situational data about the bloggers whose posts are included in the corpus I built. I am very grateful for their help. I am grateful to the Mechanical Turk workers who worked on my HITs, and I am grateful to Joe Geluso for helping me code the level of prescriptivism in usage-guide entries and for his contagious enthusiasm for research and learning.

I am grateful to Bryan Shrader for his dedicated patience as I struggled while he helped me learn the basics of Python, and I gratefully acknowledge Jesse Egbert for taking the time—when he was under no obligation—to mentor me, talk with me about my research, and build my confidence.

My classmates and colleagues deserve special acknowledgement as well. I am grateful for their friendship, support, and encouragement during my time at Iowa State.

Finally, I am forever grateful to Melissa, Day, Neve, and Kit for their unwavering love.

ABSTRACT

Since the inception of modern linguistic study in the early twentieth century, prescriptivism has largely been ignored as an area worthy of serious linguistic investigation. However, recent theoretical and empirical work has sought to better understand the relationship between prescriptivism and language variation and change. In this dissertation, I carry out a three-part study in which I compare the way that the features for eight well-known prescriptive usage rules can be observed in formal and informal writing. In Part I of the study, I identify the eight usage problems to include in the study and the 11 usage guides from which to collect relevant entries. I then present a prescriptivism index for each usage problem from each guide to represent the extent to which each guide suggests that each rule should be upheld or ignored. In Part 2, I carry out a comparative corpus analysis of the usage problems in two different registers (blogs and news) that differ in terms of formality. The purpose of this analysis is to compare the proportions of times the rules are followed or not followed between registers. In Part 3, I present the results of a survey conducted among bloggers and news writers to better understand their views on the different usage problems.

Findings revealed substantial differences in the ways that usage guides treated the usage problems and in the attitudes bloggers and journalists espoused toward the usage problems. Additionally, there was substantial variation in the degree to which individual usage rules were adhered to, but there was surprisingly little variation in the patterns of usage for each usage problem between the two registers. These findings suggest that the level of formality of the text did not considerably affect the usage patterns observed in the corpora.

CHAPTER 1. INTRODUCTION

It is no surprise that many English speakers have strong ideas about what constitutes correct and incorrect English usage. An instance of *imply* when *infer* is intended, an apostrophe placed before a plural *-s*, or the use of *literally* to mean *figuratively* are all examples of language that in the prescriptive tradition would be considered “bad” or “incorrect.” This type of usage often incites passionate criticism from self-proclaimed sticklers, mavens, snoots, and grammar Nazis—and sometimes even from everyday people who otherwise do not give much thought to language. Indeed, it seems that almost everyone enjoys the feeling of noticing and pointing out a grammatical error in one form or another.

In spite of the fact that language is constantly undergoing change, there are certain features of the language—like those given in the examples above—that certain groups of people work hard to preserve and protect. The features that draw this kind of attention are all a part of the prescriptive tradition in English. The prescriptive tradition promotes the idea that when a language contains more than one way of expressing a single meaning, one of the alternatives is correct while the others are incorrect. Though not all groups of semantically equal alternatives in a language fit into the prescriptive tradition (e.g., speakers of English can express possession with either an *-’s* or an *of*-phrase, neither of which are contested in terms of correctness) the examples above represent some well-known, though still ultimately arbitrary, rules from the prescriptive tradition: *imply* and *infer* have distinct meanings, plurals are not formed with apostrophes, and *literally* does not mean *figuratively*.

Other rules in the prescriptive tradition are less well known. One of the more obscure rules that appears in some style books and usage guides has to do with the prescriptively preferred use of the word *collide*. The *Associated Press Stylebook* (Christian, Jacobsen &

Minthorn, 2010) states that *collide* is correctly used only when it describes two objects that are in motion.¹ The entry from the 2010 *Stylebook* states it this way:

Collide, collision Two objects must be in motion before they can *collide*. A moving train cannot *collide* with a stopped train.

The 2015 edition of the *New York Times Manual of Style and Usage* (Siegal & Connolly, 2015) echoes the same rule that “[o]nly two objects in motion can collide” (p. 84) before discussing ethical issues associated with certain syntactic constructions using *collide*: “If the phrase *collided with* seems to fix blame, avoid it by using this construction: *A truck and a bus collided*” (p. 84). *Merriam-Webster’s Dictionary of English Usage* (Gilman, 1994) takes a more descriptive approach in its discussion of this usage item, citing evidence that *collide* has been used in cases where only one party is in motion since at least 1746. *Merriam-Webster* concludes its entry by assuring readers that they “will seldom have to worry about this matter” (p. 259).

Not all usage guides that include an entry on the COLLIDE rule treat the rule in the same manner. In contrast to the style manuals and guide quoted above, Garner (2016, p. 180) includes an entry for *collide* in his *Modern English Usage* but leaves out any mention of the two-items-in-motion element and instead discusses only the (in his view) appropriate prepositions to use with the term.

Collide is construed with *with* or *against*. Although *with* is more common today, the *OED* provides historical evidence of *against*, and that usage still sometimes appears—e.g.: “In the eighth he collided *against* the outfield wall while chasing a

¹ On February 21, 2018, the *AP Stylebook* dropped its entry for *collide*. The change was later officially announced at the meeting of the American Copy Editors Society in April 2018.

drive by Missouri designated hitter Jake Epstein.” Rick Cantu, “Aspito’s 3-Run Homer Saves Longhorns,” *Austin Am.-Statesman*, 20 Feb. 1999, at C3.

Interestingly, the example sentence Garner provides is one that flouts the prescription outlined in the newspaper style guides quoted above. And ironically, it is taken from a newspaper article.

The example involving the COLLIDE rule illustrates two general problems with prescriptivism in general: (1) style manuals and usage guides are inconsistent in the way they treat usage rules, and (2) some of these rules, like the COLLIDE rule just illustrated, are generally unknown and therefore ignored by everyday speakers and writers. In a 2017 post to his blog, “You Don’t Say,” John McIntyre, copy editor at the *Baltimore Sun* and former president of the American Copy Editors Society, criticized the COLLIDE rule, calling out its absurdity: “What we are dealing with here is one more rule, like the bogus *over/more than* distinction,² invented and enforced by American newspaper editors and invisible to the rest of the English-speaking world” (para. 5). He continues: “So we have literally generations of American editors who have been enforcing this preposterous rule, for which there is no legitimate basis and *which is invisible to everyone who is not a newspaper editor*” (para 13, emphasis in original).

Much of the prescriptive tradition in English is based on the idea that speakers and writers require language experts (usually not academic linguists and sometimes self-appointed) to identify and define the rules that determine what is considered good English and what is not. Some of these prescriptive usage rules are invisible in the way that McIntyre

² According to the *over/more than* rule, *over* should not be used in the sense of *more than*. For example, the sentence *The cost is over \$10* flouts the traditional rule. Prescriptivists would argue that the sentence should be revised to say *The cost is more than \$10*.

sees the COLLIDE rule, namely, they are rules that only trained professional newspaper editors know and enforce. Others (e.g., the rule to not end a sentence with a preposition) are widely known among literate speakers of the language; however, even these well-known rules are not part of the language that native speakers naturally acquire. Bourdieu (1991) highlights this fact when he describes the concept of a prescriptive usage rule as a

correct, i.e. corrected, expression [that] owes the essential part of its social properties to the fact that it can be produced only by speakers possessing practical mastery of scholarly rules, explicitly constituted by a process of codification and expressly inculcated through pedagogic work. (p. 61)

As Bourdieu argues, prescriptive usage rules are not acquired naturally; they must be explicitly taught.

Many of the prescriptive rules of English taught in schools around the United States are flouted by average, educated language users, as evidenced by the fact that many of the proscribed forms (e.g., using *me* for *I* in *Me and James are playing baseball*) are highly frequent in actual language use. Language users break the rules often in naturally occurring speech and in their informal writing, and in some cases, writers break these rules even in formal, carefully edited writing. This distinction between formal and informal writing is sometimes addressed in usage-guide entries, in which the author of the usage guide states explicitly that following a rule is expected in formal writing but not in informal writing. Yet in these entries it is often not clear what exactly the author means by “formal” and “informal” writing,³ and, despite “a greater emphasis on providing evidence of usage” in guides

³ One guide that stands as a notable exception to this statement is the *American Heritage Guide to Contemporary Usage and Style* (2005). In its front matter, Joseph P. Pickett, the guide’s executive editor, provides a brief but clear description of formal English and informal English in order to help readers understand “circumstances under which a given usage will be appropriate” (p. xiv).

published since the late twentieth century (Straaijer, 2018, p. 21), authors still (a) seldom provide specific empirical data to support their recommendations or (b) do not always make clear whether the empirical data they do offer has influenced the advice they provide. By basing their usage advice on empirical data, usage-guide authors are able to provide authoritative guidance that goes beyond simply proscribing their own personal linguistic pet peeves. Advice based on empirical data also invites readers to critically consider the advice given to determine how relevant and useful it might be. When combined with information about the rhetorical situation of a communication event such as the audience, exigence, and constraints (Bitzer, 1968), empirical data offers valuable information that can allow speakers, writers, and editors to make more fully informed decisions about the language they use (or, in the case of editors, enforce) in order to maximize the rhetorical effect of their messages. I discuss the implications that usage advice based on empirical data can have for technical editing in Section 6.2.

The current study critically and empirically considers the use of features associated with eight commonly discussed prescriptive usage rules as they can be observed in two different registers of writing: one formal and one informal. The concepts of register (along with the related concept of style) and formality are discussed in Section 1.1 below in order to provide a theoretical foundation for the study. Rather than focusing on obscure usage rules—those that are known only to trained newspaper editors, like the COLLIDE rule discussed above—the present study investigates features associated with some of the most widely promoted prescriptive usage rules to determine how their use differs across various contexts of writing. More specifically, this study offers insights into questions such as the following:

- How do usage guides differ in the ways they discuss some of the most well-known usage rules? Do different usage guides suggest that some rules are still worth following while others are not?
- Are some usage rules followed differently in different contexts? If so, which ones and to what extent?
- How do different groups of writers perceive these rules in terms of how strictly they should be followed in different contexts?

In response to the questions presented above (as well as the formal research questions presented in Section 3.4), the primary goals of this study are

- a) to study the advice given in a range of current usage guides to determine the extent to which they suggest a set of eight well-known rules should or should not be followed,
- b) to collect and analyze features associated with each of the usage rules from two different written registers—one formal and one informal—to determine the extent to which the rules are followed and not followed in both, and
- c) to survey writers of formal and informal texts in order to collect their opinions regarding the acceptability of breaking the rules in certain contexts.

Because it seeks to study linguistic variation in contexts in which more than one linguistic form can be used to express the same idea, the current study is situated in the variationist paradigm. Research in the variationist paradigm investigates variation between “alternate ways of saying ‘the same’ thing” (Labov, 1972, p. 188, quoted in Szmrecsanyi, 2019, p. 76). Szmrecsanyi continues his definition of this paradigm, saying that “variationist linguists carefully account for competing variants and draw on quantitative methodologies to model the conditioning factors that regulate the way language users choose between semantically

and functionally equivalent variants” (p. 76). In the current study, I account for variants within individual usage problems that essentially allow users multiple ways of saying the same thing; however, only one of these ways is generally regarded as prescriptively correct. I use the register of the discourse, comparing two registers that differ in terms of formality, to observe whether register constitutes a “conditioning [factor] that regulate[s] the way language users choose between semantically and functionally equivalent variants” as Szmrecsanyi noted.

The overarching goal of the study is to offer empirical evidence that can inform future discussions about the usage rules under investigation here, including how writers and editors might better determine when and when not to follow or enforce the rules in different contexts.

1.1 Register, Style, and Formality: A Theoretical Framework

The concepts of register, style, and formality provide important theoretical foundations for studying prescriptive usage problems. Register and style are related concepts⁴ in that both allow analysts to map lexico-grammatical features to various functional and aesthetic attributes of a text. Register and style are related to formality because both can be considered more or less formal or informal. Because formality plays a large role in discussions of prescriptive usage problems, it is important to define and explain this term. In this section, I offer an overview of each concept and discuss the reasons they are important in studies of prescriptive usage problems.

⁴ Making meaningful distinctions between these terms has been a problem for researchers for decades. Lee (2001) likens distinguishing between these related concepts to “walk[ing] into a well-known quagmire” (p. 41), a quagmire which I now find myself in!

1.1.1 Register

Registers are commonly conceptualized as language varieties that are defined by their situational contexts, not by the linguistic similarities or differences they may exhibit (see Biber & Conrad, 2009; Gray, 2015, ch. 4). A register perspective of language assumes that all language is functional, responding to the requirements of the situations in which it is produced. In this view, the reason certain groups of texts (i.e., registers) show similar systematic linguistic patterns is because these linguistic patterns are required by the situation (Biber & Conrad, 2009, p. 18). As an illustrative example, let us consider the register of academic journal articles. The texts that make up this register are produced in different situations from other written registers, such as fiction novels. Both are written, revised, and edited, but because academic articles are written for a highly specialized audience who share a substantial amount of background knowledge with the author of the text, and because academic journal articles must often adhere to strict word limits, they tend to use linguistic features like complex noun phrases and technical jargon that allow writers to compress a lot of information into shorter texts (see Biber & Gray, 2016). These linguistic features are functional because they respond to an exigence in the situation.

In many conceptions of register, the situation is primary, meaning that language production occurs inside an already existing situation; it does not *create* the situation. Biber and Conrad (2009) noted that “the situational characteristics of registers are more basic than the linguistic features” (p. 9). They elaborate further, saying

All speakers use language in different contexts, under different circumstances, for different purposes. Those patterns of behavior cannot be derived from any linguistic

phenomena. In contrast, the linguistic differences among registers can be derived from situational differences, because linguistic features are functional. (p. 9)

Bitzer (1968) made a similar point half a century earlier in his definition of the rhetorical situation, saying that “[r]hetorical discourse is called into existence by situation” (p. 8). Rhetorical situations, according to Bitzer, must include an exigence, an audience, and constraints. The exigence, according to Bitzer, is “an imperfection marked by urgency” (p. 6) and can be considered rhetorical when it can be changed or altered through discourse. The audience is the group of people who are “mediators of change” (p. 7). The constraints are the “persons, events, objects, and relations” (p. 8) that are needed to change the exigence. The definition of a sociolinguistic situation put forth by Brown and Fraser (1979) overlaps to some degree with Bitzer’s definition of a rhetorical situation described above. The sociolinguistic situation comprises a setting (analogous to Bitzer’s constraints), purpose (analogous to Bitzer’s exigence), and participants (analogous to Bitzer’s audience). Biber and Conrad (2009) provided perhaps the most comprehensive account of situational characteristics. Their framework for analyzing situational characteristics of a register includes identifying the following situational features about the texts in the register:

- the participants involved in creating the texts (e.g., how many there are, power differences among them, the amount of knowledge they share, etc.)
- the mode and circumstances of production (e.g., whether the texts are produced spontaneously or revised and carefully edited)
- the setting (e.g., the time and place in which the texts are produced)
- the purpose the texts (e.g., to inform, narrate, explain, etc.)
- the topic of the texts (e.g., academics, lifestyle, politics, etc.)

Situations are composed of smaller situations, even down to the level of individual speech acts. Brown and Fraser (1979) use the example of a visit to the doctor to demonstrate how an individual situation can comprise many smaller situations. For instance, when a person visits the doctor, they⁵ may drive to the doctor's office, check in with the receptionist, and wait in the waiting room before actually being seen by the doctor. Each of these events leading up to the actual doctor exam can be considered individual situations that combine to form the macrosituation of the doctor visit. Taking the example one step further, each of these subsituations might also include additional microlevel situations, analyzable on the speech-act level. For instance, asking the receptionist a question might be considered a different situation from listening to the response.

Recognizing the situation as primary in a register framework is critical because it is only through comparing the situational characteristics of two or more registers that any linguistic differences observed can be noticed and meaningfully analyzed (Biber & Conrad, 2009, p. 36). Because situations can be analyzed on different levels of detail, register analyses can be conducted on different levels as well. Biber's (1988) foundational register analysis investigated variation between spoken and written registers generally by considering a range of more specific registers within speech and writing. More recent analyses have investigated variation in registers that are much more closely related in terms of their situational characteristics. For example, Gray's (2015) study of variation in academic writing took the work of previous scholars who compared registers based mostly on the situational variable of academic discipline (e.g., biology or political science) and extended it to also consider the methodology employed in the articles (i.e., quantitative vs. qualitative) as an

⁵ In this dissertation, I use *they* as a gender-neutral singular pronoun.

important situational variable. Gray found that the method used was a significant factor in determining linguistic variation across registers. Thus, registers and their situational contexts can be defined in a near infinite number of ways.

1.1.2 Style

Biber and Conrad (2009) note that the concepts of register and style are similar, in that they can both be used to “[analyze] the use of core linguistic features that are distributed throughout text samples from a variety” (p. 2) but that they differ with respect to the motivations that underlie the variation in linguistic features. Namely, a register perspective notes functionally motivated differences (i.e., features that can be observed as a result of differences in the situational contexts in which different texts were produced) while a style perspective notes aesthetically motivated differences. Lee (2001) highlights the idea that style is best conceptualized at the level of individual text because it “characterise[s] the internal [i.e., linguistic] properties of individual texts or the language use by individual authors” (p. 45). Register, according to Lee, is a concept that views the language of texts (note the plural) in terms of their situational characteristics. In this respect, register is a broader concept than style (Biber & Finegan, 1994, p. 5) because it exists across texts where style exists within individual texts.

It is important to distinguish between *stylistics* on the one hand and *style as a concept within register analysis* on the other. Stylistics is a “method of textual interpretation” (Simpson, 2004, p. 2) that involves “looking systematically at the formal features of a text in determining their functional significance for the interpretation of the text in question” (Jeffries & McIntyre, 2010, p. 1). Similarly, Verdonk (2002) called stylistics “the analysis of distinctive expression in language and the description of its purpose and effect” (p. 4). In

both Jeffries and McIntyre's and Verdonk's definitions, stylistics can be used to uncover the function of linguistic features in a text. In contrast, linguistic style as it is conceived in register analysis is not viewed as functional because it does not respond to the situational characteristics in which the text was produced. In the register perspective, style serves only aesthetic purposes (Biber & Conrad, 2009, p. 18).

These different views of the role of style are noteworthy because they both acknowledge the important role that situation plays in language use, but they differ in the situations they emphasize. In register analyses, the situation refers to the context in which the text was *produced*; in stylistics, the situation refers to the context in which the text is *consumed*. While considering style does not allow functional determinations to be made based on the situation of production, it can lead to useful observations about the situation in which the text is consumed and interpreted.

In studies of prescriptive usage rules, register and style are important concepts because usage problems represent linguistic alternations that are both functional (register) *and* aesthetic (style). Usage problems serve the social purpose of differentiating between different levels of social status based on educational differences (see Ebner, 2017, p. 7). Therefore, the choice to follow one linguistic alternation over another represents a functional choice—one constrained by the setting, purpose, and participants (Brown & Fraser, 1979) of a given situation. Usage problems are also aesthetic because they represent what influential prescriptive linguists—and many times the authors of texts as well—consider to be the most clear, concise, and logical language, that is, the language that is most aesthetically pleasing.

Consider the well-known usage problem of the split infinitive as an illustrative example of how usage problems can be viewed as both functional and aesthetic. When an

author writes a clause that includes a verb in the infinitive form modified in some way by an adverb, the author has the choice to place the adverb between the two components of the infinitive verb as in (1) or to place the adverb somewhere else in the clause as in (2).

(1) She wants to further study the issue.

(2) She wants to study the issue further.

This linguistic choice is at its core stylistic. The placement of an adverbial in a sentence is not necessarily functionally motivated by the situational characteristics in which the text is produced but instead represents an aesthetic difference that some may argue is based on clarity or logic. But because this particular type of adverbial placement—split infinitives—has been marked as socially taboo for centuries, the choice to split the infinitive or not also has functional relevance because the author’s choice to do it or not can depend on the situation in which it the feature is produced. A situation in which a writer knows their audience will harshly judge people who flout the rule must take this situational characteristic into account before determining whether to split the infinitive or not. In this way, usage problems can be seen as both aesthetically and functionally motivated.

1.1.3 Formality

I have noted in the previous two sections the ways in which register and style have been commonly conceived in previous scholarship, including some of the ways in which the two concepts are similar and different. One important commonality these two concepts share is the fact that they can both be analyzed in terms of formality. We can talk of “formal style” and “formal registers.” Formal styles are created when authors use formal linguistic features; formal registers are produced in formal situations, such as those that call for seriousness, politeness, or respect (Irvine, 1979). In this section, I describe in fuller detail the different

ways that formality has been conceptualized as existing at the linguistic level (style) and at the situational level (register).

The distinction between formal and informal contexts has been an important foundational distinction in sociolinguistic work, though it is seldom clearly defined. Irvine's (1979) analysis of literature in sociolinguistics and related fields revealed that the concepts of formality and informality are often used in a way that lacks analytical clarity. She found that some research seemed to situate formality as originating within the linguistic features of a text, while other literature seemed to view formality as originating in the social situations in which language is produced. These opposing views of formality can be seen in more recent research. Heylighen and Dewaele (1999) take the first approach in their definition of formality, determining the level of formality of a text based on its linguistic characteristics. They define formal language as language that is "context independent and precise" and that "represents a clear distinction which is invariant under changes of context" (p. 8). Thus, they argue, texts that include many context-independent linguistic features (e.g., nouns, adjectives, prepositions, and articles) will result in a text that is more formal, while texts that contain many context-dependent linguistic features (e.g., pronouns, verbs, adverbs, and interjections) will result in a text that is more informal. The features identified in Heylighen and Dewaele's work align consistently with Biber's (1988) distinction between informational and involved content and serve, essentially, as a simplified version of the first factor in Biber's important multi-dimensional analysis of speech and writing—a factor that shows a "fundamental dimension of linguistic variation among texts" (see pp. 102–108) and one that has been viewed as marking a distinction between formal and informal variation (e.g., Grieve, Biber, Friginal & Nekrasova, 2011, p. 309).

Empirical work has also been done to determine formality at the lexical level. Levin, Long, and Schaffer (1981) and Levin and Novak (1991) conducted experiments to determine whether participants perceived Latinate or Anglo-Saxon words to be more formal. These studies found that participants generally identified the Latinate words as more formal than their Anglo-Saxon synonyms. In Levin and Novak's study, the participants were not provided information about the context or the situations of the utterances they were asked to evaluate, and so their opinions of formality were based solely on the words themselves. More recently, Thayer, Evans, McBride, Queen, and Spyridakis (2010) asked participants to rate the level of formality of different versions of online texts that differed with respect to seven lexico-grammatical features: absence or presence of *I*, *we*, and *you*; use of active or passive voice; absence or presence of verb contractions; use of formal versus informal punctuation; and absence or presence of the word *welcome*. They found that all seven features had a significant influence on the ways that participants perceived the formality of the texts, and conclude that "certain syntactic and semantic choices influence the perception of a text's formality" (p. 456).

Studies such as those just discussed offer evidence that formality may exist at the lexico-grammatical level. Of course, it is possible that participants associate lexico-grammatical features such as Latinate words with formal situations and that is the reason that they identify them as more formal than their Anglo-Saxon counterparts. In this way, the relationship between formality and level of language might be conceptualized as a back-and-forth relationship with an indeterminate beginning. In other words, formal situations can necessitate formal linguistic features, but formal linguistic features may also be able to influence the level of formality of a given situation (see Levin & Novak, 1991, p. 390). From

a register framework, however, situation is primary, meaning that situations determine linguistic features, but linguistic features do not determine situations, though they can be markers of formality (Biber, 1988, p. 33; Brown & Fraser, 1979).

In addition to conceptions of formality at the linguistic level, formality has also been defined at the situational level. Trudgill (1999) characterized styles in terms of formality, stating that they “can be arranged on a continuum ranging from very formal to very informal” with “[f]ormal styles...employed in social situations which are formal, and informal styles...employed in social situations which are informal” (p. 119). Since Labov’s (1966) pioneering study of social stratification in New York City, the field of sociolinguistics has placed a crucial emphasis on studying casual linguistic style or “the every-day speech used in informal situations, where no attention is directed to language” (p. 100). Sociolinguists “seek to study ordinary speech in everyday life” (Baugh, 2001, p. 112) and therefore place a premium on creating comfortable, informal situations in which they can elicit casual, natural language to study in their research. Eliciting casual speech is important to sociolinguists because it helps them mitigate what Labov (1972) called “the observer’s paradox” (p. 209), or the way that speakers tend to use guarded, formal speech when they know their speech is an object of study—in other words, when they feel they are participating in a formal situation. In this way, sociolinguists have historically been interested in studying informal speech (seldom writing), and they have determined the informal nature of the speech they studied based on the situations in which the speech was produced.

Similar to the views of sociolinguists just summarized, scholars who study register variation have also defined formal/informal language in terms of the situation in which the language is used. As noted above, registers are defined by their situational characteristics and

are thus inherently linked to the situational contexts in which they occur. Finegan and Biber (1994, p. 316) note that early studies of sociolinguistic variation (e.g., Labov, 1972) looked only at attention paid to speech as the influencing factor of variation. A register perspective, in contrast, is multifaceted and can thus account for a fuller range of observed variation in language. That is, in addition to the attention paid to language production, register variation considers the participants, their relationship with each other, the purpose of communication, and the topic of communication (among other variables) as factors that affect variation. Each of these factors can be analyzed on a scale of formality. For instance, a communicative event in which two close friends discuss their plans for the weekend over lunch is likely to be much less formal than a state dinner between the president of the United States⁶ and the prime minister of England. The fact that the friends know one another well and are engaged in a discussion about an informal topic in an informal setting all influence the informality of the speech that is likely to take place. On the other hand, the exchange between the two powerful leaders, acting in their respective capacities as heads of state, could include discussion of consequential topics and would almost certainly be heavily scrutinized by the media and citizens of their countries, thus creating a highly formal setting in which a formal exchange is to be expected. Viewing linguistic variation from a register perspective is thus a useful approach because it allows researchers to conduct a fuller analysis of factors that influence the formality of a given register than previous work in sociolinguistics allowed.

Because the current project adopts a register framework, I adopt the view of linguistic formality as (a) being determined by the situation in which the language is used, not by the

⁶ Because the current president of the United States so often flouts the conventions long established for the office of the presidency and formality in general, I would ask that the reader imagine almost any other past president when considering this example.

individual language features produced, and (b) existing on a cline ranging from informal to formal.

In the current study, I identify two different registers, personal blogs and news writing, as occupying different places on the formality cline in order to compare the degree to which a set of well-known usage rules are or are not followed in each. I argue that personal blogs are comparatively less formal than news writing because they are not formally associated with professional institutions and are thus able to discuss a wide variety of topics from serious to frivolous. News writing on the other hand is highly constrained and associated with media institutions that seek to maintain a professional ethos. In this way, I argue news writing is more formal than blog writing. I provide a fuller situational analysis of these two registers in Chapter 3

Studying variation in these two registers is valuable because much of the usage advice offered in the prescriptive tradition is couched in terms of the formality of the situation in which a text is produced (see Section 2.5), yet no studies of which I am aware have done a careful situational analysis to determine the level of formality of one register over the other and then gathered and analyzed empirical data to determine whether the rules are followed more or less in formal or informal registers. This study is an attempt to begin filling this gap in the literature.

1.2 Contribution of the Study

This study demonstrates the ways that prescriptive language rules are of interest to scholars from technical communication and linguistics. Technical communication is a humanistic discipline (Miller, 1979) grounded in a rhetorical tradition that is concerned with social justice and egalitarianism. For technical-communication scholars, then, the current

study highlights and analyzes the self-sustaining social structure that allows people in power to create the rules that define “good English,” and it offers empirical data that scholars, teachers, and practitioners can use to make better-informed decisions regarding the language they use. For linguists, the study contributes to a growing interest in linguistic scholarship that advocates for investigating the ways that prescriptivism is related to language variation and change. Through this study, I hope to add to the work that others have done in building a bridge between the fields of technical communication and applied linguistics, and I hope that the results of this study will help to contribute to discussions of the degree to which language use varies across formal and informal registers—particularly with respect to prescriptive usage items.

The study has important practical implications practice as well. Prescriptive usage problems are commonly addressed in writing classrooms and ESL curricula, and they are a near-constant concern for technical editors. Through empirically analyzing usage patterns of the usage problems under investigation, the results of this project can help writers, editors, and language learners make more informed, native-like choices when considering the language they use; it can help writing and editing teachers help their students understand that not all prescriptive language rules contained in textbooks and style manuals need to be observed in all cases; and it can help professional editors make informed decisions about which usage rules they will enforce and which they will not.

1.3 Overview of the Study

In the following chapter, I review the literature associated with prescriptivism and its role as a subject worthy of serious study. In Chapter 3, I offer an analysis of the situational characteristics of the two registers I have selected to represent formal and informal writing

and present the three research questions this dissertation seeks to answer. In Chapter 4, I discuss the proposed research design and method of analysis of the study. Chapter 5 contains the results of the analyses I conducted of the usage problems included in this study (comprising Sections 5.1–5.8). For each usage problem, I review the usage-guide authors’ advice and present the overall prescriptivism index. I then describe the specific methods used to analyze the proportions in which each rule is followed in both registers, and I present the results of that analysis. I offer a summary of the results from the survey I conducted as described in Section 4.3. Finally, I synthesize these findings in Section 5.9. In Chapter 6, I conclude by reviewing the findings of each research question, identifying limitations, and discussing the implications this study—and others like it—might have for technical editors.

CHAPTER 2. LITERATURE REVIEW

2.1 Prescriptivism as an Area of Linguistic Interest

In a 1953 essay published in *College English*, Morton Bloomfield advocated for the value of prescriptivism in the teaching of English. At the time of the essay's publication, descriptive linguistics was quickly becoming mainstream and Bloomfield was an early critic, pointing out the potential pedagogical follies that come with adopting a completely descriptive view of language. Bloomfield framed his argument in terms of facts and values, emphasizing the idea that questions of pedagogy should take both into consideration. Bloomfield acknowledged the importance of the facts (i.e., the observations about language made by descriptive linguists), but suggested that values (i.e., the social weight that proper English carries) should not be lost in the mix.

Ultimately, the question which we must basically consider in dealing with the teaching of English is what kind of men [sic] we want to make of our students. This cannot be solved by a knowledge of the history of the English language. (p. 34)

In this way, Bloomfield argued that values should be prioritized over linguistic facts in language pedagogy.

Almost half a century later, Curzan (2002) made similar arguments for students to understand prescriptive language, pointing out that

It is possible to teach Standard English [i.e., the prescriptions that make up Standard English –JS] while at the same time creating a meta-awareness of that educational process, so that students are empowered to examine the system and its language hierarchies critically, so that they can challenge that view if they should choose to—with full control of the language variety of power. (p. 342)

While Curzan did not advocate for teaching Standard English as strongly as Bloomfield did, she did recognize the need to teach the *politics* of Standard English because it can have important effects on the social status of students and it can help them consider issues related to Standard English critically.

The ideology that Bloomfield feared—the one that prized facts about language over values—is the ideology espoused by most academic linguists. These scholars disparage prescriptive attitudes toward language use, opting instead to adopt the more scientific approach of describing language variation without placing value judgments on these variations. McWhorter (1998) succinctly summarized how he views the harmful effects of prescriptive ideologies on grammar in this way: “Prescriptive grammar has spread linguistic insecurity like a plague among English speakers for centuries, numbs us to the aesthetic richness of non-standard speech, and distracts us from attending to genuine issues of linguistic style in writing” (p. 62). Because of its harmful effects, as many linguists see them, prescriptivism has largely been ignored or shunned in linguistic research. Milroy and Milroy (1999) cataloged some of the dominant attitudes on the topic, stating that many linguists (Milroy and Milroy *not* included) see prescriptivism as “irrelevant to linguistics” (p. 4). Chapman described the common conception of prescriptivism among linguists as “the ugly stepchild in the scholarly community” (personal communication, October 17, 2017), as evidenced by the fact that there is not yet an association devoted to the study of prescriptivism’s influence on language, there is no journal dedicated to publishing articles about prescriptivism, and conferences centered on prescriptivism are still sparsely attended. Universal prescriptive usage rules are even challenged in applied fields (Boettger & Wulff,

2014; Connatster, 2004; Mackiewicz, 2003), demonstrating that the legitimacy of many of these rules is widely contested.

One reason linguists and others may bristle at the concept of prescriptivism lies in the fact that prescriptive imperatives are not arguments; that is, they do not seek to gain their authority through evidence or systematic inquiry. Chapman (2010), making reference to etiquette manuals, the forerunner of usage guides (discussed in Section 2.5), likened prescriptive rules to those rules that dictate polite manners, calling them part of a canon in which “their tenets are inherited and received, rather than questioned and proved” (p. 142). Earlier, Cameron (1996) pointed to a similar characteristic of prescriptive edicts, saying that “linguistic imperatives are really moral imperatives” because “the ‘force’ of a linguistic prescription has little to do with persuasion in the sense of rational argument” (p. 8) and is instead based on appeals to speakers’ inherent sense of right and wrong. Linguistics as a discipline views language from a scientific standpoint and as such is primarily interested in evidence and argument. It is not surprising then that such an uncritical view of language would seem groundless to anyone interested in the serious study of language.

Despite these prevailing attitudes, scholarship devoted to the serious study of prescriptivism has been steadily increasing among linguists—primarily those interested in discourse analysis, the history of English, and sociolinguistics—over the last several decades. Milroy and Milroy (1985/1999) were perhaps the first to undertake a serious book-length study of the relationship between prescriptivism and language variation and change. In their study, they argued that “[a]lthough it is necessary to insist on the priority of description, it does not follow from this that prescription should never be studied at any point” (p. 4). They see prescriptivism as an integral part of language standardization, occurring after codification

(the emergence of grammars and dictionaries of a language) and thus as the final stage in the implementation of a standard language. Cameron's (1995) heavily cited work took the concept of prescriptivism beyond the notion of enforcing rules codified in grammars and dictionaries and extended it to the broader concept of what she calls verbal hygiene, or "an urge to 'clean up' language" (p. 1). In doing so, she pointed out that the popularly anti-prescriptivist approach to language taken up by most (if not all) academic linguists—an approach in which natural, spontaneous variation is seen as good while efforts to encourage language users to consciously change their language are deemed bad—is itself a prescriptive ideology and that it "in a certain sense...mirrors the very same value-laden attitudes it seems to be criticizing" (p. 3) because it suggests that there is a right way to think about language change and a wrong way to think about language change. Where other commentators have likened English to an irrepressible natural phenomenon with implications that it changes on its own (see Curzan's [2014] analogy of living language as a river), Cameron underlined the fact that language is inherently a social force used by humans for human communication. This idea is important when considering verbal hygiene—and prescriptive language ideologies in general—because "[a]rguments about language use are not between man and nature, they are between groups of people with differing opinions and interests. It is not 'the English language' that verbal hygiene attempts to subdue, it is particular users of that language" (pp. 22–23).

2.2 Prescriptivism as a Social Phenomenon

The sociological view of language that Cameron (1995) and others adopt offers a sound theoretical foundation for studying prescriptive language ideologies. Bourdieu (1991) wrote extensively about institutions and processes that must be in place in order to maintain

the “legitimate language,” which he described as “a semi-artificial language which has to be sustained by a permanent effort of correction, a task which falls both to institutions specially designed for this purpose, and to individual speakers” (p. 60). Because the legitimate, or “correct,” language is semiartificial, it is not naturally self-sustaining, and thus requires people and institutions to constantly correct those who do not yet speak the legitimate language. Some of those who are corrected (particularly those who are corrected as children or during their schooling) eventually become the next generation of correctors and in this way create a self-sustaining social structure that gives them the power to define and prescribe “good English.” Bourdieu referred to it as a “process of continuous creation” in which various authorities compete “for the monopolistic power to impose the legitimate mode of expression” and thus “ensure the permanence of the legitimate language and of its value, that is, of the recognition accorded to it. (p. 58).

The social issues that many Western linguists see in prescriptive ideologies are not necessarily found in the prescriptive traditions of all languages. For instance, Myhill (2004a) contrasted American views of prescriptivism with Israeli views, arguing that American linguists view prescriptivism negatively because they conflate two different reasons for the problem of prescriptivism, namely, prescriptivism represents ideologies that are both societally unfair and scientifically flawed. However, Myhill argues that not all prescriptive attitudes toward language include both reasons. Standard, prescriptively correct Hebrew, for example, represents a variety that is scientifically unsound, but because it is based on appeals to ancient religious texts and not on the usage patterns of elite Hebrew speakers, prescriptivism in Israeli Hebrew is not socially unfair. This means that “Hebrew speakers and linguists interpret these criticisms not as further arguments that prescriptivism is evil and

should be done away with altogether, but rather as showing that there are limitations on what can be expected of it” (p. 30). In a related paper, Myhill (2004b) argued that the concept of correctness should be reconceptualized to reflect these different social views. He identified three different types of correctness:

1. Textual correctness: based on writing
2. Prestige-based correctness: based on the usage of the social elite
3. Prescriptive correctness: based on the declarations of a recognized authority (such as a language academy)

All three concepts of correctness are scientifically unsound, but as Myhill argued, so are many social conventions. Simply because they are scientifically unsound does not offer reason enough to discount them completely. However, only prestige-based correctness has the problem of privileging one social class above another in an unfair way, and for this reason, it needs to be rethought. In contrast, Venter (2002) highlighted the social nature of a correct Standard English, fully embracing the social distinctions that Myhill argued are so problematic:

I believe we should celebrate the wonderful diversity of our dialect—or our dialects. This does not, however, mean that there is no standard, but that the standard must be upheld for the right reasons, and in the right way. The reason is social, not linguistic. (p. 71)

In spite of arguments for or against it, the reality is that prescriptivism in American English is fraught with social problems. Even though language continues to change in spite of the social structures in place and continued efforts of correction, these social structures still have important effects on the way language is used and on how language is viewed in

society. Recently, Curzan (2014) warned that “histories of English are mistaken if they minimize or marginalize the modern prescriptive project as failed because it has failed *to stop* those alterations which time and change have made in language” (p. 3). She goes on to argue that “[i]n marginalizing prescriptivism, [scholars] can miss important developments in Modern English usage and in meta-discourses about usage, both of which are part of language history” (p. 3). In other words, continuing to ignore prescriptivism as an object worthy of serious linguistic attention will result in an incomplete understanding of English and will not do anything to bring together the divide between linguists, prescriptivists, and the general public. (See Tieken-Boon van Ostade, 2013, for a brief description of the Bridging the Unbridgeable project, which, as its name implies, seeks to bridge this divide.)

2.3 Origins of Prescriptivism

Peters (2006) noted that prescriptive approaches to English began to emerge in the eighteenth century as “a byproduct of eighteenth-century efforts to codify the grammar of the language” (p. 761). Unlike most of the contemporary major grammars of modern English, the grammars of the eighteenth century contained prescriptive usage advice in order to improve what Lowth (1799), in his widely influential grammar, called “grammatical accuracy”:

The English language hath been much cultivated during the last two hundred years. It hath been considerably polished and refined; its bounds have been greatly enlarged; its energy, variety, richness, and elegance, have been abundantly proved, by numberless trials, in verse and in prose, upon all subjects, and in every kind of style: but, whatever other improvements it may have received, it hath made no advances in grammatical accuracy. (p. iii)

Lowth's own *A Short Introduction to English Grammar*, the work from which the above quotation comes, marked an important moment in the history of English. Millward (1996) called Lowth's work "the most prominent" of all the eighteenth-century grammars and noted that "[m]any of his [Lowth's] decisions about English usage have come down to us virtually unchanged" (p. 244). Indeed, as Millward pointed out, it was Lowth who first proscribed the use of double negatives and insisted that words like *only* and other adverbials be placed next to the words they modify in a sentence. To Lowth and other usage experts of his time, the idea of grammatical accuracy was important, in part, because they believed

that there existed a 'universal' grammar from which contemporary languages had deteriorated. Greek and Latin were (wrongly) assumed to have deviated less from this original purity than had the various European vernaculars, and thus they (especially Latin) were regarded as models upon which an improved English grammar should be based. (Millward, 1996, p. 242)

It is likely that fewer people today view language as inherently divine, but the forcefulness with which many people today still adhere to the prescriptions that originated during this time of intense focus on the English language more than 200 years ago still influences the way many people think about "good English."

2.4 Prescriptivism Today

Today, major grammars of English (e.g., Biber et al., 1999; Huddleston and Pullum, 2002, Quirk et al., 1985) are largely descriptive, though some smaller grammars do treat grammar from a prescriptive framework (e.g., Parker & Riley, 2005). For prescriptive commentary on English usage, one must now turn to usage guides, a genre of language-related texts that seek to help readers make sense of different kinds of usage problems or

contested lexical and grammatical issues. Stamper (2014) noted the influence of a changing social structure in the rise of usage guides. According to Stamper, English speakers in the late 1600s held a generally negative view of their own language. They did not simply believe that *other* people spoke English incorrectly; they believed that the language itself was wrong—to the point that even the poet laureate of the time could not write poems in English. The language was considered vulgar—particularly for polite society. Over time, these same associations of illegitimacy spread until they were not only applied to the English language, but to the speakers of English as well. People reasoned that if a speaker had been speaking a “wrong” language all their life, they must have absorbed all the wrong rules of language in general. The upward social movement of the middle class corresponded to this heightened awareness of language usage, such that by the 1700s, middle-class men (and they were all men) were looking for ways to adjust to their newly improved position in society. One way of doing so was to ensure that they spoke as correctly as possible. Etiquette guides that taught these men about politeness, genteel behavior, and the proper use of language became best-sellers. These etiquette guides constitute the first usage guides for native speakers of English.

The desire to speak and write correctly remains a concern for a large number of people, as evidenced by the increasing numbers of books that are printed each year on the subject. This is likely fueled by the continued feelings of personal linguistic insecurity that still abound even in speakers who feel that their regional variety is correct (Preston, 2013, p. 322). Miłkowski (2013) pointed out that interest in correct usage is not limited to prescriptive purists only, but extends to a broader sample of language users (p. 175), further emphasizing the wide-reaching effects that prescriptive language ideologies have on people’s attitudes toward language use.

Even though the serious study of prescriptive rules in English has generally received scant attention in linguistic research, issues related to prescriptivism provide interesting and fruitful areas of study from a sociolinguistic perspective. The study of usage problems (i.e., linguistic variations in which one variant is coded as correct and the others are coded as incorrect) fits well within the purview of sociolinguistics—or more specifically, the sociolinguistic subfield of language variation and change—because the usage problems themselves represent variation in the language.

2.5 Usage Problems and Usage Guides

Usage problems are discussed in usage guides, reference materials that often give advice about when and when not to use the different variants. These discussions are often framed in terms of formality. Usage-guide authors may advise readers to adhere to a prescription only in formal situations, though what exactly constitutes a formal situation is a topic that is rarely clearly defined. For example, *The Longman Guide to English Usage* (Greenbaum & Whitcut, 1988) advises readers that “split infinitives should...be avoided in formal writing whenever possible” (p. 671) but does not comment on what type of writing readers should consider to be “formal.” Because so much writing that people currently produce can be considered “an everyday social practice” that “involves both strong and weak regulation depending on the context and ways in which [it] is gatekeepered” (Lillis, 2013, p. 173), empirical studies that distinguish between formal and informal written registers are becoming increasingly important. And because prescriptive rules are often couched in discussions of formal and informal contexts, they provide a useful area for better understanding the differences between formal and informal writing.

Usage guides as a genre can be challenging to define. Straaijer (2018) noted that a usage guide “lists the meanings of words, but is not a dictionary.... [It] discusses grammatical structures, but is not a grammar” (pp. 11–12). Weiner observed that the purpose of usage guides is to offer readers “a short cut to the acquisition of habits” that are “acquired, not automatically—through growing up among speakers of the language—but through a conscious educational process” (Weiner, 1988, p. 172). Hundreds of usage guides are available today, and many are widely read. Pullum (2010) noted that Strunk and White’s *Elements of Style* has sold over 10 million copies and that “[m]any college-educated Americans revere *Elements*, swear by it, carry it around with them” (p. 34) despite the fact that the book’s success is, in his opinion, “one of the worst things to have happened to English language education in America in the past century” (p. 34). Lynne Truss’s *Eats, Shoots, and Leaves* (2004) has even topped best-seller lists. Indeed, many native English speakers today believe in a correct way to speak the language and even feel a sense of self-consciousness that their own language somehow does not live up to that standard. Instead of feeling that the language is wrong, as those in the seventeenth century did as noted above, many speakers today believe when a speaker has violated a rule contained in a usage guide, it is the *speaker* who is wrong (see Milroy, 1992, pp. 31–32; Pullum, 2004, p. 3), and not some problem with the language itself. Usage guides, thus, provide a means by which speakers and writers might improve their own English or at least bring it into conformity with some standard they see as desirable.

Usage guides are composed of an author’s comments on a series of usage problems (see Ilson, 1985, for an early use and discussion of the term). A usage problem is a specific type of linguistic alternation variable. Grieve (2016) called a linguistic alternation variable “a

set of two or more ways of saying the same thing” (p. 38). A usage problem, then, is a set in which one of the ways of saying the same thing is contested in terms of its correctness. For instance Sentences (3) and (4) are functionally and semantically equivalent; however, Sentence (3) is considered incorrect because it breaks the rule outlined in many usage guides that proscribes the use of *fewer* as a modifier for a plural countable noun—in spite of the fact that *less* occurs with plural countable nouns even in standard edited English.⁷

(3) *I’ll take less items if it means I don’t have to carry as much.

(4) I’ll take fewer items if it means I don’t have to carry as much.

Ebner (2017) emphasized that usage problems are socially constructed and serve the divisive function of differentiating between correct/educated language use from incorrect/uneducated language use (p. 7).

It is helpful to think of usage problems as comprising two primary components: a traditional rule (the prescription), and an authentic, attested usage that flouts the prescription. Chapman (2010) argued that usage problems must be widespread enough that they are noticeable but not so widespread that they are considered unobjectionable; in his words, “[h]aving a prescription attract notice takes just the right amount of exposure of the construction” (p. 143). Follett (1966) took the idea of “attract[ing] notice” further, extending it to actually *caring* about the distinction inherent in the usage problem: “A linguistic pattern [i.e., usage problem] is dead, not when there is a large amount of deviation from it...but when it has ceased to make a clear and uniform impression upon those who attend to words

⁷ A simple search for *less* + [plural noun] and *fewer* + [plural noun] in the Corpus of Contemporary American English (Davies, 2008) shows that, while the construction with *fewer* is substantially more frequent in written registers of standard edited American English, the proscribed construction with *less* still occurs. The results of these searches revealed the following frequencies per million words in four different registers: Fiction: 0.78 (*less*), 4.04 (*fewer*); magazine writing: 1.65 (*less*), 23.08 (*fewer*); newspaper writing: 1.91 (*less*), 24.58 (*fewer*); academic prose: 1.77 (*less*), 25.21 (*fewer*). The frequencies for the constructions with *less* are not substantially large; nevertheless, they do demonstrate that *less* is used to modify countable nouns in edited American English.

(p. 373). In other words, a usage problem is dead only when people “who attend to words” no longer care about the issue.

2.6 The Interaction between Prescriptivism and Usage

This brief overview of usage problems, usage guides, grammar writing, and the rise of prescriptivism is general and far from complete. But it sets into relief how far back beliefs about what is and what is not “good English” go. The historical duration of these attitudes underscores the importance of the recent calls for a serious study of prescriptivism. Scholars like Curzan (2014), Cameron (1995), and Milroy and Milroy (1999), as noted above, have done important work in establishing a theoretical grounding for studies in prescriptivism. Recent empirical work (e.g., Anderwald, 2016; Auer & González-Díaz, 2005) has taken up these scholars’ call to action by investigating the interaction between prescriptive ideologies and changes in actual language use. Auer and González-Díaz (2005) found mixed results in their study of prescriptive grammaticography’s⁸ influence on language change. The authors conducted a case study using corpus methods to study two usage problems: the inflected subjunctive (*if I were able to attend* versus *if I was able to attend*) and the double comparative (*more better*). They found evidence to suggest that prescriptive writing in the eighteenth century may have had some effect on the use of the inflected subjunctive: grammarians at the time lamented its loss, and subsequently, its use began to increase. The influence of grammar writing on the double comparative was harder to discern, however. The authors found that the use of the double comparative had at the time already been reduced to near zero by the time the grammar writers composed their works in which they oppose it, so

⁸ Grammaticography is generally understood as the study of grammar writing. Anderwald (2016) suggested that the study of grammaticography can be used to answer questions such as “Which terminology was used, which definitions, and with which examples? Which phenomena were explained in detail, and which only rarely? What is the common core of...grammar writing?” (p. 9).

their comments on the usage problem “can only be considered a mere reinforcing factor of a process that was set in motion much earlier” (p. 336). As such, the cause of the decline in usage cannot be tied to the grammar writers’ comments on the topic.

In her study of nineteenth century grammaticography, Anderwald (2016) compared prescriptions and proscriptions contained in 258 nineteenth-century grammars with historical corpus data to “[correlate] what happened in language...with what the grammars had to say” (p. 9). She found that the grammars written during this century took a “generally prescriptive stance” (p. 237), but that the prescriptions and proscriptions found in the grammars often did not precede similar changes in actual language use. In fact, Anderwald found that, in many instances, the reverse was true: as the language continued to change, the grammars followed suit. For example, where the majority of American grammars published at the beginning of the nineteenth century insisted on *shrunk* as the only acceptable past tense form of the verb *shrink*, more than 80% of American grammars published at the end of the century either preferred *shrank* or insisted on *shrank* as the only correct form, a trend that she observed in historical American English usage (p. 66). Thus, Anderwald provided important evidence of the ways that prescriptive grammars and actual usage interact with one another.

It is important to point out that it is difficult—if not impossible—to state that prescriptive rules *cause* some change in the language. Curzan (2014) rightly stated that “[t]he interaction of prescriptivism and usage defies straightforward cause–effect relationships” (p. 87). However, when considered as a whole, research that observes the interaction of prescriptive rules and changes in actual usage, Curzan argued, “make it clear that prescriptivism should not be dismissed as *a factor* in the development of formal written English” (p. 84, emphasis mine). In other words, we can conclude with some level of

confidence that prescriptivism has at least *some* effect on language variation and change—particularly in written English, as Curzan stated above—but the amount of this effect is not able to be discerned through correlational corpus studies alone. However, studies that investigate the relationship between prescriptivism and actual language use can still offer potentially useful probable explanations for some of the influences that cause changes in language to take place.

As the above studies show, research into prescriptivism's influence on actual language use yields somewhat conflicting results and consensus is far from being reached. Much additional work remains to be done to fully understand the potential ways in which artifacts of prescriptive ideologies can be observed in everyday language—particularly the English used in the twentieth and twenty-first centuries. The current study contributes to the growing interest in the interaction between prescriptive language ideologies and actual language use by offering comparisons of the ways prescriptive rules are and are not followed in different registers of writing. Unlike some of the studies mentioned earlier, the current study will not take a historical perspective but will instead identify the extent to which prescriptive usage advice can be observed in two different registers of Present Day English: formal written English (news writing) and informal written English (personal blogs). In the next chapter, I discuss the ways that these two registers differ in terms of formality, specifically noting that because news writing is authoritative and often carefully edited before publication to ensure it adheres to some standards of language use, it is more formal than personal blog writing—a register that is usually not seen as authoritative or as carefully edited according to formal language guidelines before it is published. Following this analysis

of the situational characteristics of these two registers, I present the research questions guiding this study.

CHAPTER 3. REGISTER ANALYSIS

In Chapter 1, I proposed a working definition of linguistic formality that clearly positions it within the situational characteristics of a communicative event. In the current chapter, I discuss the ways in which the situational characteristics of news writing and personal blog writing differ and thus contribute to my argument that they represent comparatively formal and informal writing, respectively.

3.1 Justifying the Choice of Personal Blogs and News

There are, no doubt, many registers that one could choose to represent formal and informal writing. Academic articles, legal documents, and technical documentation might all be argued to represent formal writing. Similarly, text messages, social media posts, and online forum discussions might all be argued to represent informal writing. In this study, I have selected personal blogs to represent informal writing and news to represent formal writing. The primary reason I selected personal blogs as a register of informal writing was that, as Crystal (2006) noted, blogs offer a unique opportunity to collect extended samples of discourse that have not been edited in the way that other more formal registers of writing have been. The extended nature of the discourse is important in order to build a corpus that is large, but still principally sampled. Many personal blogs are also freely accessible and not hosted behind firewalls or other obstacles to collecting the texts, so it is feasible to build such a corpus, as I detail in the next chapter. Of course, simply collecting a large sample of blogs of any kind would pose methodological issues for this analysis, as blogs in general cannot usefully be considered a register in the same way that books cannot be usefully considered a register. Both are publication types (Egbert, personal communication, January 19, 2018), meaning that they provide a vehicle for delivering content to an audience, but the content

published in them can vary so widely that considering the situational characteristics that all books or all blogs have in common would be so general as to make a register analysis of that kind less than valuable. For this reason, I have selected to include only personal blogs in this analysis in an attempt to narrow the register I work with in this project. Personal blogs have been identified as a register in previous studies (e.g., Biber & Egbert, 2018) and are more situationally unified than considering blogs in general. I discuss the process of collecting the blogs used for this study in Chapter 4.

The primary reasons I selected news as a register of formal writing is twofold. First, Biber (1988, p. 37) identifies “informational exposition” (i.e., news writing) as representing typical or unmarked writing. Because written registers are commonly viewed as being more formal in nature than spoken or other oral registers, the “typical or unmarked” nature of news makes it a useful register to represent the formal nature of written discourse in general. Certainly more formal registers of writing exist (e.g., legal discourse, academic writing); however, these registers are not as familiar to general audiences and might therefore be considered hyperformal. For the purposes of this study, I preferred to analyze a register that could reasonably be considered formal while still having widespread readership. The second reason I selected news as a register to represent formal writing is that well-designed corpora of news writing already exist and are freely available, making a comparison of news writing relatively convenient. In this study, I study news writing using the news subcorpus of COCA (Davies, 2008).

Biber (1988) noted that “[d]escribing the situation is a precursor to functional descriptions of language use” because it “enables an interpretation of the roles played by particular linguistic features within that context” (p. 33). According to Biber, analysis of

situational characteristics allows researchers to “link the functions of particular linguistic features to variation in the communicative situation” (p. 33). This section describes the situational characteristics of the two registers I investigate in this study. I begin by comparing the similarities and differences in the situations in which news writing and personal blogs are produced. I then turn to a more in-depth discussion of the ways that the situational characteristics of news writing have changed in recent decades and how even in spite of these changes, news writing can still be viewed as a comparatively formal register. I conclude with the research questions that guide this study.

3.2 Comparing the Situational Formality of Personal Blogs and News

Myers (2010) contended that “[b]logs take up some of the functions of newspapers” but do so “with a more personal perspective and less institutional weight” (p. 11). Indeed, some blogs are used for the same communicative purposes found in news writing: namely, to describe information and narrate or report on events. In addition, both news writing and blog writing are consumed online (often, in the case of news writing; always, in the case of blog writing), and both are composed of many, sometimes unrelated, discrete elements—articles and columns in news writing and posts in blog writing—so readers can easily consume small portions of information from either source. Both are also intended for a public audience, though the size and reach of news writing is in many cases much larger than that of blog writing.

Aside from these similarities, there are also many situational differences related to formality between these two registers. For example, personal blog posts can be written about any topic for nearly any purpose. News articles, on the other hand, are written only about topics that editorial boards consider newsworthy for the purpose of informing the public. As

a result, news writing is always timely and intended to be considered important and authoritative while blog writing can sometimes be frivolous and at times even silly or senseless. Because of these different situational constraints, news writing is almost always more formal than blog writing.

Another important way that blog writing and news writing differ situationally is the fact that blog writing is more intrinsically linked to the author of the content than news writing is. Hyland and Feng (2017) argued that *informality* in writing

is not simply a reluctance to attend to conventional practice, any more than it is an inappropriately colloquial use of language. Instead it is an attempt to establish a particular kind of relationship with readers, one which makes assumptions about a shared context and seeks to create a collegial familiarity. (p. 42)

Blog writing accomplishes this level of familiarity—and thus informality—through the author’s sharing of personal stories, anecdotes, and other details about their lives, thus fundamentally linking the blog author to their content. Myers (2010) described this feeling of familiarity by saying, “Blogs speak to a group [as opposed to an individual], but a group that seems to know each other and share a lot already. One sinks into a blog as one sinks into the sofa in a friend’s living room” (p. 93). This same level of familiarity is not observed in news writing (with the possible exception of opinion articles or weekly columns). In general, news writing is highly informational. The content is presented as factual reporting with only the reporter’s byline to provide information about the author of the article. Thus, for most readers of news writing, the writing is completely disassociated from its author, which contributes to its formal nature.

While both news writing and blog writing are produced in similar circumstances—in which authors are able to draft and revise their original work before publishing it to their audiences—the level of care (Mackiewicz, 2010) and attention paid to language (Labov, 1966; 1972) is often higher in news writing. Even in spite of changes to the amount of editing that news writing undergoes (as I discuss in Section 3.3), the amount of revision and editing done for each register still likely differs to some degree and thus constitutes an important situational difference between blog writing and news writing for the purposes of my study.⁹ Owen (2013) found that *educated* writing is not the same as *edited* writing and thus the two cannot be conflated. The fact that news writing is often carefully edited before publication with trained copy editors ensuring the articles meet language standards (see Bell, 1991, pp. 75, 82) contributes to the formal nature of news writing. Crystal (2006) argued that because blog writing is not carefully edited in the way news writing is, it presents the “opportunity to see written discourses of sometimes substantial length which have had no such editorial influence” and is therefore “written language in its most naked form” (p. 245). As a result, blog writing—even when written by educated native English speakers—is a valuable resource for studying informal features of language.

3.3 Changes in the Production of News Writing

Previously, all credible news writing underwent rigorous copy editing. Indeed, Bell (1991) noted that “copy written by one newswriter is always edited by another” (p. 75). It is easy to assume that news writing could be considered formal primarily because it is carefully

⁹ In survey data collected for this purposes of this study, however, a higher proportion of bloggers (74.03%) said that they carefully edited their blog posts as they write or after they write. Only 65.38% of journalists noted the same thing. While this data is not representative of all bloggers and journalists and therefore cannot be generalized to the entire population, it does prompt interesting questions about the ways that blog writing and news writing are produced. I return to this issue in Sections 5.9.1 and 6.1.3.

edited. However, recent changes in the way news writing is produced have affected the ways in which news writing is reviewed and edited. Declining revenues and increasing competition from alternative news sources have contributed to substantial losses in revenue for many print news sources. By some estimates, revenues from advertisements dropped by two-thirds just in the period from 2000 to 2015 (Perry, 2015). These severe reductions in profitability have forced newspaper companies to consolidate, reduce, or in some cases eliminate copy-editing roles (see Jarvis, 2012; Keith, 2015). As a result of these factors—as well as news reporters operating within a 24-hour online news cycle with pressure to almost instantly publish stories once they are written—news copy today as a whole is not as carefully edited as it once was. This does not mean that all news writing is published without careful review, however. Major newspapers such as *The Baltimore Sun* still employ copy editors who review news copy before it is published (*Baltimore Sun*, 2018), and major “digitally native” news sources such as vox.com (Vox, n.d.) employ dedicated copy editors who presumably ensure the copy in the articles they publish is clear and accurate. However, one overall effect of declining revenue for newspapers has been a general reduction in the attention paid to copy editing in news writing.

While the staff cuts and consolidations summarized above do complicate any claim that news writing is carefully and thoroughly edited, it does not completely discredit the argument that news writing represents a formal written register of English. Recent research has shown that rates of mechanical errors in news writing have remained constant over a roughly 80-year period (Beede & Mulnix, 2017), suggesting that even in the wake of major reductions in newsroom staff in the early 2000s, the grammatical correctness of news writing in general has not suffered. Additionally, because news reporters are trained writers, they are

professionalized in a way that other writers (including many blog authors) are not. Many reporters working for major news outlets almost certainly have degrees in journalism or have undergone some other form of formal training in news writing (the reporters I personally know offer anecdotal evidence of this claim). In addition to being trained writers, journalists working as news writers are expected to have some knowledge of copy editing. Spanogle (2014) states that “[i]n high school journalism labs...all members of the staff become journalists learning copy-editing skills” (p. 49), and LaRocque (2012) reminded reporters that they are responsible for finding and fixing their own errors of “language basics—grammar, spelling, punctuation” (p. 9). Thus, the training that reporters have is an important situational characteristic to consider in the situational analysis of news writing and contributes to its formal nature.

Having formal education and training in a field is a necessary step in becoming professionalized in that field. For professional communicators (a group to which news writers certainly belong), professionalization of this kind goes beyond the practical skills news writers acquire during their training; it also allows them to develop a deeper sense of the social and ethical responsibilities that come with their job as a news writer. In analyzing the ways that professional identities are characterized in academic literature, Faber (2002) found that

Professionals have an integral relationship with a specific and known audience, they have unique social responsibilities that extend beyond the workplace and into general society, and they have a self-conscious, self-reflexive ethical awareness that simultaneously creates and protects their occupational identity as sole experts in their field. (p. 316)

These deeply held relationships, responsibilities, and feelings of self-awareness contribute to the professional identity of news writers and place constraints on the type of work they produce. In a way, it makes their work more serious. Blog writing, in contrast, is not a professionalized skill and as a result is perhaps less likely to be influenced by the kinds of relationships, responsibilities, and self-awareness Faber described. Though some people are indeed “professional bloggers,” and there are some professional organizations that bloggers can belong to, there is no requirement that a person must have formal training in blog writing to become a blogger. That is, there is no formal, recognized process of professionalization that a blogger must complete in order to begin blogging. Certainly, the vast majority of people who keep a personal blog do not belong to a professional organization, nor do they likely have a formal degree in blog writing (if such a thing exists—a quick internet search seems to suggest it does not). As a result, blog writers, unlike news writers, have the freedom to write with more flexibility due to the lack of expectations and constraints that their situation imposes. In this way, the situation in which blog writing is produced is more casual and therefore informal (see Irvine, 1979). This casual nature of blog writing is one characteristic that attracts readers to blogs. It allows bloggers to mimic the casual ways that friends converse and emphasizes underlying commonalities that bloggers and their readers share (see Myers, 2010, p. 93). The professionalized situation in which news writing is produced, in contrast, places many constraints on the news writers, and therefore results in a more formal situation of production, one that emphasizes the authority of the news organization and their role to inform rather than entertain readers.

3.4 Research Questions

Developing a full analysis of the situational differences between news writing and blog writing is an important step in building a convincing argument that these two registers are in fact different in terms of the level of formality that can be observed between them. The situational analysis presented here has highlighted some of the key differences in the situations in which news writing and blog writing are produced, focusing specifically on the professional(ized) role that news writers take on when they produce their writing and contrasting it with the lack of training required to produce blog writing. In this way, I argue that news writing and blog writing can be productively used to represent written registers on different ends of the formality continuum.

Formality is a topic that is addressed in many usage guides. In their discussions of specific usage problems, these guides sometimes suggest certain rules need not be followed in all contexts; sometimes the (in)formality of the situation is given as a possible factor to consider when determining whether the rule should be followed or not. For this reason, it will be important to first study the entries for the usage problems included in this study to determine the extent to which they suggest that these rules should be followed. Because of these differences in terms of formality, I expect to observe a higher proportion of adherence to prescriptive usage advice in the more formal register of news writing than the more informal personal blog writing I study. It will also be instructive to ask writers of blogs and news articles their own opinions about these usage problems to determine whether they find breaking the rule to be unacceptable, acceptable only in certain contexts, or acceptable in any context. With these three considerations in mind, the specific research questions the current study seeks to answer are the following:

- RQ1 To what extent do American usage guides recommend following the traditional rules for eight well-known usage problems?
- RQ2 To what degree do formal edited writing (news) and informal unedited writing (personal blogs) differ in the ways they conform to the prescriptive usage advice given for the same set of eight usage problems?
- RQ3 Do bloggers and news writers/editors perceive these usage problems differently? If so, to what extent and in what contexts?

In the next chapter, I outline the research design I have developed to answer these questions.

CHAPTER 4. RESEARCH DESIGN

To investigate the three research questions posed above, I adopted an explanatory mixed-methods research design (Creswell et al., 2008, as cited in Mackey & Gass, 2016), in which “qualitative data are collected after the collection of quantitative data and are used to help explain the quantitative results” (pp. 281–282). In this section, I describe the method of collecting and analyzing data for each of the research questions presented above.

4.1 Determining the Level of Prescriptivism in Usage Guides (RQ1)

The first research question asked the extent to which current usage guides recommend following the traditional rules for eight well-known usage problems. The purpose of this research question is to provide a baseline that can help interpret the results from the corpus analysis, introduced in Research Question 2 above and described in detail in Section 4.2 below. In other words, in order to more fully understand why one rule may or may not be followed in different registers of writing, it will first be necessary to understand whether current usage guides recommend that their readers follow the rule in the first place. For example, if most current usage guides recommend that their readers can confidently end their sentences with prepositions (flouting the well-known rule that prohibits this practice), then that would offer valuable information in interpreting the extent to which that rule could be observed in different written registers. Of course, it would be irresponsible to suggest that the level of prescriptivism in the guides has a direct effect on the actual usage observed in the corpora; however, knowing how prescriptively each usage problem is treated in the guides can still be helpful for interpreting the findings observed in response to Research Question 2.

Determining the level of prescriptivism for each guide included five general steps: (1) selecting the usage problems to include in the study, (2) selecting the usage guides to include

in the study, (3) compiling the relevant entries from the guides into a rating sheet, (4) developing a scale to calculate a prescriptivism index, and (5) using the rating scale to calculate a prescriptivism index for each usage problem in each guide. Each of these steps is treated in its own individual section below.

4.1.1 Selecting the Usage Problems to Include in the Study

The first step in determining the level of prescriptivism was to identify a set of specific usage problems to include in the study. I used the Hyper Usage Guide of English (HUGE) (Straaijer, 2015) to identify usage problems that are discussed in a large number of American usage guides. The HUGE database contains 77 usage guides (44 of which are published for an American, as opposed to British, audience) and catalogs a total of 123 usage problems. Thus, the HUGE database allows users to easily compare which usage problems are contained in different usage guides and, where copyright permissions have been obtained, to read the entries for these usage problems from the usage guides.

To select which usage problems to include in this study, I first identified the 10 usage problems that appear in the highest number of American usage guides in the HUGE database. I then selected eight that could be feasibly analyzed from a corpus-based perspective. Table 4.1 shows the 10 usage problems that appear in the highest number of American usage guides in the HUGE database. The shaded rows indicate usage problems that I did not include in this study.

Table 4.1 Usage problems included in the current study. The terms are taken from the HUGE database and the example sentences are adapted from those offered by the HUGE database.

Usage problem	Abbreviated description of the rule	Example sentence where traditional rule is flouted. (The prescriptively preferred form is shown in parentheses below.)	No. of American usage guides that discuss this usage problem
SHALL/WILL	Use shall in first-person constructions to express futurity and in second- and third-person constructions to express volition. Use will in complementary distribution.	*I <u>will</u> be 21 tomorrow. (I <u>shall</u> be 21 tomorrow.)	37
LAY/LIE	Use lay only as a transitive verb and use lie only as an intransitive verb.	*She had the tools just <u>laying</u> around. (She had the tools just <u>lying</u> around.)	37
WHO/WHOM	Use who in subject position always and use whom in object position always.	* <u>Who</u> did he ask? (<u>Whom</u> did he ask?)	35
DIFFERENT TO/THAN/FROM	Different from is correct. Different to is incorrect. Different than is correct only when what follows is a complete clause.	*Running is very <u>different than</u> jogging. (Running is very <u>different from</u> jogging.)	36
SPLIT INFINITIVE	Splitting infinitives is never correct.	*She refused <u>to even think</u> of it. (She refused <u>even to think</u> of it.)	33
ONLY	only must be placed immediately next to the word/words it modifies.	*He <u>only had</u> one chapter to finish. (He <u>had only</u> one chapter to finish.)	34
I FOR ME	Use object pronouns in object position and subject pronouns in subject position.	*She told <u>Charles and I</u> the whole story. (She told <u>Charles and me</u> the whole story.)	34
SINGULAR THEY	they is a plural pronoun and therefore cannot be used with a singular antecedent.	* <u>Everyone</u> has <u>their</u> own style. (<u>Everyone</u> has <u>his or her</u> own style.)	35
LESS/FEWER	Use less to modify noncountable nouns and use fewer to modify countable nouns.	*There were <u>less accidents</u> this year than last. (There were <u>fewer accidents</u> this year than last.)	35
NONE IN PLURAL CONTEXT	None is always singular and should therefore agree with singular verbs.	* <u>None were</u> left on the table. (<u>None was</u> left on the table.)	32

Note: Shaded rows indicate usage problems that I did not include in this study.

As shown in Table 4.1, the usage problems I eliminated are SHALL/WILL and ONLY. After conducting a pilot analysis of the SHALL/WILL usage problem, I determined that it is likely too problematic to pursue in the current study because *shall* has been steadily declining in frequency (Leech, Hundt, Mair & Smith, 2009, p. 79) and it appeared in the pilot analysis mostly in quoted text, which may be reflective of its declining use. Because of this, I was not able to make meaningful comparisons between the use of *shall* and *will*. I eliminated the ONLY usage problem because the research methods that would be required to properly analyze each instance fall outside the research design presented here. In order to accurately assess whether each instance of *only* was used according to the prescriptive rule or not, I would need to assess the writer's intended meaning (i.e., whether their placement of *only* in the clause accurately conveyed the meaning they intended or not). To do this would require interviews with authors about specific instances of this usage problem as it appears in their writing, which was outside the scope of the study presented here. Studying eight usage problems still allowed me to analyze a wide range of usage problems while still being feasible for a study of this size.

4.1.2 Selecting the Usage Guides to Include in the Study

As described in the previous section, the eight usage problems I chose to include in the study are discussed in as many as 37 usage guides in the HUGE database. Carefully studying the entries for all eight usage problems in every usage guide in which they are discussed was not feasible for the current study, as it would have resulted in hundreds—possibly more than 1,000—pages of material to analyze. As a result, it was necessary to select a sample of the usage guides to include in the study. To do this, I identified the 10 most current usage guides included in the HUGE database. The compilers of the database

prioritized collecting first editions of the usage guides; however, because I am interested in *current* attitudes toward usage, I prioritized collecting the most current editions of the guides. Therefore, if a guide in that list had a more recent edition than the one included in the HUGE database, I used the entries from the more current version. I also included the most current editions of *Merriam-Webster's Dictionary of English Usage* (Gilman, 1994) and Strunk and White's *Elements of Style* (2009) even though these guides did not appear on the list of the 10 most current guides contained in the HUGE database. I included these additional guides because of their popularity and influence. The most current guides included in the HUGE database and the editions I used in the present study are listed in Table 4.2

4.1.3 Compiling the Relevant Entries from the Usage Guides into a Rating Sheet

After I had selected the eight usage problems and 11 usage guides to include in the study, I compiled all of the entries for each of the usage problems from all of the guides listed in Table 4.2 into a single Word document in order to facilitate more efficient evaluation. I included in the document what I determined to be the most important entries from the usage guides that corresponded to each usage problem. In many cases, a usage guide addressed a single usage problem in more than one entry. For example, Garner (2016) addresses the I FOR ME usage problem under entries titles “B. Between you and me; *between you and I,” “CLASS DISTINCTIONS,” “HYPERCORRECTION,” and “PRONOUNS.” In this case, I included the relevant text from each of these entries in the rating sheet in the section for this usage problem.

Table 4.2 Comparison of usage guides in the HUGE database and those included in this study

Guides in HUGE database				Editions used in this study				Notes
<i>Author</i>	<i>Title</i>	<i>Ed</i>	<i>Year</i>	<i>Author</i>	<i>Title</i>	<i>Ed</i>	<i>Year</i>	
Fogarty, M.	Grammar Girl's Quick and Dirty Tips for Better Writing	1	2008	Fogarty, M.	Grammar Girl's Quick and Dirty Tips for Better Writing	1	2008	--
Peters, P.	The Cambridge Guide to English Usage	1	2006	Peters, P.	The Cambridge Guide to English Usage	1	2004	The edition included in the HUGE database is the second impression, which is why the year of publication is different between the one in the database and the one included in this study.
--	The American Heritage Guide to Contemporary Usage and Style	1	2005	--	The American Heritage Guide to Contemporary Usage and Style	1	2005	--
Bakto, A.	When Bad Grammar Happens to Good People: How to Avoid Common Errors in English	1	2004	Bakto, A.	When Bad Grammar Happens to Good People: How to Avoid Common Errors in English	1	2004	--
Brians, P.	Common Errors in English Usage	1	2003	Brians, P.	Common Errors in English Usage	3	2013	--

Table 4.2 continued

Trask, R. L.	Mind the Gaffe: The Penguin Guide to Common Errors in English	1	2001	Trask, R. L.	Mind the Gaffe: The Penguin Guide to Common Errors in English	1	2006	Originally published in the UK as <i>Mind the Gaffe: The Penguin Guide to Common Errors in English</i> , this guide was previously published in the US as <i>Say What You Mean!</i> in 2005. <i>The New Fowler's Modern English Usage</i> (2000) is the third edition (second impression—the first was published in 1996) of the original <i>Fowler's</i> (1926) and it was edited by Robert Burchfield. The <i>Pocket Fowler's Modern English Usage</i> (1999) is a condensed version of the third edition. It was edited by Robert Allen (<i>A Dictionary of Modern English Usage</i> , n.d.). Even though the two are based on the same original 1926 guide, they are edited by different people, which is presumably why both were included in the HUGE database. However, both books have new editions. The new edition of <i>The New Fowler's Modern English Usage</i> is called <i>Fowler's Dictionary of Modern English Usage</i> (4th edition) and the new edition of the <i>Pocket Fowler's Modern English Usage</i> is called <i>Fowler's Concise Dictionary of Modern English Usage</i> (3rd edition). Both are edited by Jeremy Butterfield. Since both new editions are based on the same guide and both are edited by the same person, I included only <i>Fowler's Dictionary of Modern English Usage</i> (4th edition) in my data.
Fowler, H. W. and Burchfield, R.	The New Fowler's Modern English Usage	3	2000	Butterfield, J.	Fowler's Dictionary of Modern English Usage	4	2015	
Fowler, H. W. and Burchfield, R.	Pocket Fowler's Modern English Usage	1	1999					

Table 4.2 continued

Garner, B. A.	A Dictionary of Modern American Usage	1	1998	Garner, B. A.	Garner's Modern English Usage	4	2016	--
O'Conner, P. T.	Woe is I	1	1998	O'Conner, P. T.	Woe is I	3	2009	--
--	Webster's Dictionary of English Usage	1	1989	--	Merriam-Webster's Dictionary of English Usage	2	1994	--
Strunk, W.	The Elements of Style	1	1918	Strunk, W. and White, E. B.	The Elements of Style	4	2009	This anniversary edition is a reprint of the fourth edition with an added note from the publisher.

I used the entries cataloged in the HUGE database as a starting point for determining which entries to include in the rating sheet. In some cases, I included entries that were not cataloged in the HUGE database—possibly because they were simply missed or they may not have appeared in an earlier edition of the usage guide. In compiling the rating sheet, I took care to include the entries that dealt substantively with the usage problems under investigation in this study. However, including every section that referenced each usage problem in some way would have caused the rating sheet to become prohibitively long (as used in the study, it already contained circa 60,000 words). For example, the SINGULAR THEY usage problem is discussed in entries for individual indefinite pronouns (e.g., *anyone*, *anybody*, *someone*, *somebody*, etc.) in some usage guides because singular *they* is often used in anaphoric reference to indefinite pronouns. It would have been prohibitive to include each entry for every indefinite pronoun from these guides.

Each entry included in the rating sheet was carefully compared with the text in the published guides to check for accurate transfer—both in terms of content and format (i.e., italicized words remained italicized). I also made some changes to the formatting of some entries, e.g., replacing spaces with tabs, indenting example sentences to make them easier to identify, and adding some descriptive information to some of the abbreviations included in the guides. A portion of the rating sheet is shown in Figure 4.1.

ID	USAGE PROBLEM	GUIDE	YEAR	ENTRY	RATING	NOTES (optional)
92	different to/than/from	Common Errors in English Usage_2013	2013	<p>Common Errors in English Usage_2013</p> <p>58 (83 in 2013)</p> <p>different than/different from/different to</p> <p>Americans say “Scuba-diving is different from snorkeling,” the British often say “different to” (though most UK style guides disapprove) and many say “different than,” though to some of us this sounds weird. However, though certain conservatives object, you can usually get away with “different than” if a full clause follows: “Your pashmina shawl looks different than it used to since the cat slept on it.”</p>		
117	different to/than/from	Fowler’s Dictionary of Modern English Usage_2015	2015	<p>Fowler’s Dictionary of Modern English Usage_2015</p> <p>212–213 (216–217 in 2015)</p> <p>different. 1 The commonly expressed view that <i>different</i> should only be followed by <i>from</i> and never by <i>to</i> or <i>than</i> is completely untenable in the face of past and present evidence and of logic. Fowler wrote in 1926 that insistence ‘that <i>different</i> can only be followed by <i>from</i> and not by <i>to</i> is a superstition’. It is, however, a superstition that refuses to lie down and die. (a) <i>Current use.</i> The OED dates the earliest examples of each construction as <i>different from</i> 1590, <i>different</i> to 1526, <i>different than</i> 1644. In all the varieties of English the OEC covers, <i>different from</i> is by far the most usual, followed by <i>different than</i>, then <i>different to</i>. Both British and American English show a marked preference for <i>different from</i>; <i>different to</i> is widely used in British and Australian English, and uncommon in American English; <i>different than</i> is used frequently in North America but rarely in Britain.</p> <p>(b) <i>Different from</i> versus <i>different to</i>. <i>Different to</i> is questioned on the premise that we do not say <i>differ to</i>. By this argument, all words in the same morphological family should be used with the same prepositions; e.g. we ought to say <i>according with</i> (instead of <i>according to</i>) because we say <i>accords with</i>. Other word families where the preposition changes include <i>full of</i> as against <i>filled with</i>; <i>proud of</i> as against <i>pride in</i>. But in all kinds of circumstances, <i>different to</i> is perceived as a perfectly valid alternative, perhaps also influenced by the analogy of <i>indifferent to</i>, rather than <i>indifferent from</i> or <i>than</i>, and by antonyms of <i>different</i>, such as <i>comparable to</i>, <i>equivalent to</i>, and <i>similar to</i>. Examples: (from) <i>Writing for the web is different from any other kind of writing</i>—OEC, 2002 (AmE); <i>He’s no different from my other brother, in the end</i>—N. Gordimer, 1988 (S.Afr.); (to) (rarer in AmE) <i>Road to Perdition is vastly different to Mendes’ engaging debut American Beauty</i>—OEC, 2002 (BrE); <i>Geoffrey Palmer is a vastly different prime minister to his predecessor</i>—B. Jesson in <i>Metro</i> (NZ), 1990.</p> <p>(c) <i>Different than</i>. As previously mentioned, this is a predominantly North American construction, and is often used when <i>different</i> affects a subsequent clause: <i>Farming is different than teaching in that farmers rarely congregate, especially farmers from different countries</i>—OEC, 2004 (AmE); <i>i was really no different than any other guy in my District</i>—OEC, 2005 (Canad.); (with subsequent clause) <i>It was an unequal fight, but the outcome was different than one would have expected</i>—OEC, 2002 (AmE); <i>It was in so many ways entirely different than he could ever have imagined</i>—Internet website, 2004 (AmE).</p> <p>There is a strong argument that <i>different than</i> is a useful way of avoiding cumbersome wordiness when what follows is a relative clause or a prepositional phrase. Using <i>different from</i> in the last sentence above would entail rewriting it as <i>entirely different from what he could ever have imagined</i>. To take another example, <i>the American theatre, which is suffering from a different malaise than ours</i>, is greatly preferable to <i>suffering from a different malaise from that which ours is suffering from</i>. In fact, the use of <i>different than</i> here makes it possible to summarize the whole clause from that <i>which ours is suffering from</i> in the two words <i>than ours</i>. Economy is everything. In the sentence <i>this year conditions are slightly different than in the past</i>, <i>than</i> similarly stands in for the verb (<i>than they were in the past</i>), and has the additional merit of avoiding the unwieldy combination of prepositions that <i>different from in the past</i> and</p>		

Figure 4.1 Excerpt of the rating instrument used to rate the level of prescriptivism in usage guides

4.1.4 Developing a Scale to Calculate a Prescriptivism Index

In order to rate the level of prescriptivism of each usage problem, I developed a scale that could be used to calculate what I call a prescriptivism index for each usage problem in each usage guide. Previous researchers have developed frameworks for evaluating prescriptive grammars, writing manuals, and usage guides. Peters and Young (1997) studied 11 usage problems in 40 usage guides and determined whether the entries in each text treated the traditionally dispreferred variant in each case as unacceptable, acceptable, or usable in restricted contexts. Mackiewicz (1999) developed a framework to gauge prescriptivism in writing handbooks that distinguished between prescriptive, semiprescriptive, and prescription-breaking entries, and she assigned a prescriptiveness calculation to each of the handbooks she studied. More recently, Anderwald (2016) assessed the ways various nineteenth-century grammars discussed a number of usage problems associated with verbs by

noting whether the grammars in her study mentioned one or both variants or stated a preference for one or the other (see pp. 72–73). Yáñez-Bouza (2015) used a three-part coding scheme to classify the ways that eighteenth-century grammars and rhetorics discussed the still-commonly-heard rule that writers should not end sentences with prepositions. She coded each comment about preposition stranding in the books she studied as either critic (i.e., criticizing the practice of preposition stranding), advocate (i.e., advocating for the use of preposition stranding), or neutral (p. 30). Finally, Ebner (2017) used a “slightly modified” (p. 416) version of Yáñez-Bouza’s framework for coding the entries in the usage guides included in her study by classifying them as either (a) criticizing the contested usage, (b) taking a neutral stance toward the usage, or (c) advocating for the contested usage.

While all of the frameworks described above have their strengths, I developed a different framework that allowed for a slightly more nuanced analysis of the usage guides in my study. I opted to use a modified version of Garner’s (2016) scale for assessing the acceptability of usage problems (see Garner, 2016; Peters, 2018). However, I used the scale not to assess the usage problems, but the level of prescriptivism in the entries of the usage guides. Unlike Yáñez-Bouza (2015) who coded each comment in the grammars and rhetorics she studied, I assigned an overall value to each usage guide based on the entries from each one that were contained in the HUGE database. (Some guides contained multiple entries dealing with the same usage problem while others contained only one entry for a usage problem.)

The scale I developed is reproduced in Figure 4.2. Ratings on the scale range from 1 (the guide rejects the rule and approves of breaking the rule in any context) to 4 (the guide suggests upholding the rule in all contexts). Levels 2 and 3 on the scale denote guides that

offer advice but acknowledge the role that context plays in a reader's choice to follow the rule or not.

<i>Prescription-Breaking</i>		<i>Prescriptive</i>	
+	-	-	+
1	2	3	4
Rejects the rule	Rejects the rule, but...	Upholds the rule, but...	Upholds the rule
Entry rejects the traditional rule.... Entry may overtly contradict the rule (Mackiewicz, 1999). Entry approves of breaking the rule in any context. Entry may suggest (implicitly or explicitly) that the traditional, prescriptive rule is no longer in force.	Entry generally rejects the traditional rule...but still recommends following the rule in certain contexts (e.g., formal writing). Entries that make a distinction based on formality or register fall into this category if they suggest that a person's default position can be to break the rule.	Entry generally upholds the traditional rule...and is predominantly prescriptive but acknowledges exceptions to the rule (i.e., contexts in which breaking the rule is acceptable). Entries that make distinctions based on formality or register fall into this category if they suggest that a person's default position should be to follow the rule.	Entry upholds the traditional rule...and is highly prescriptive. Entry suggests that readers should follow the rule in all contexts. Entries in which the rule is simply stated with no explicit acknowledgement of exceptions fall into this category. Entry may label nontraditional uses with pejorative labels (e.g., error, wrong, incorrect, etc.).

Figure 4.2 Scale created to rate the level of prescriptivism for each usage problem in each usage guide

4.1.5 Using the Rating Scale to Calculate a Prescriptivism Index for Each Usage

Problem in Each Usage Guide.

Two people (the primary researcher and a PhD candidate in Applied Linguistics) coded the data, assigning a value of 1, 2, 3, or 4 from the prescriptivism scale shown in Figure 4.2 to the content related to each usage problem from each usage guide in the rating sheet. The second coder was trained before completing the task. Training included reviewing definitions of the traditional rules for the eight usage problems, explaining the prescriptivism rating scale, reviewing examples of coded entries, and completing a practice coding session that included discussing our responses.

It is important to state clearly that we did not code each individual entry from the usage guides; instead, we considered all relevant entries from each individual guide together when coding each usage problem. For example, both raters assigned a single prescriptivism index to Garner's (2016) treatment of the I FOR ME usage problem and not to each of the four separate entries related to this usage problem from Garner's work that were included in the rating sheet. When assigning a prescriptivism index to the data, we read the entry (or entries, in cases where there were multiple entries from a guide for a single usage problem) and evaluated it based on the way it discussed the usage problem. We looked specifically for text in which the author suggested that the rule should be followed in every instance, only in certain instances, or if the rule could confidently be ignored. The text in the guides varied to a great degree, so evaluating them was a subjective exercise. Table 4.3 shows sample excerpts from the entries in the rating sheet along with the prescriptivism index both coders assigned to it. Text that was particularly indicative of the rating assigned is underlined in the far-right column.

The inter-rater reliability coefficient (Cohen's kappa) of the coding was 0.608, which can be considered "good" (see Fleiss, Levin, and Paik, 2003, p. 604). The percent agreement was 70.93%; the percent adjacent agreement was 94.19%. The total prescriptivism index for each usage problem was calculated by averaging the prescriptivism index from both coders. The index for each usage problem is presented in their respective sections in Chapter 5.

Table 4.3 Excerpts from the usage-guide entries that exemplify each assigned index. Each of these four examples were taken from entries to which both coders assigned the same index.

Assigned index	Guide	Usage problem	Text excerpt
1 (rejects the rule)	Peters (2004)	SINGULAR THEY	Language historians would note that the trend towards using they for both plural and singular is exactly what happened with <i>you</i> some centuries ago (see you and ye). The trend is probably “irreversible” (Burchfield, 1996). <u>Those who find it uncomfortable can take advantage of the various avoidance strategies mentioned under he and/ or she, to be used when grammatical liberties with they/them/their are unthinkable. Yet that kind of response to singular they/them/their is no longer shared by the English-speaking population at large. Writers who use singular they/them/their are not at fault.</u>
2 (Rejects the rule, but...)	Brians (2013)	DIFFERENT TO/THAN/FROM	Americans say “Scuba-diving is different from snorkeling,” the British often say “different to” (though most UK style guides disapprove) and many say “different than,” though to some of us this sounds weird. However, though certain conservatives object, you can <u>usually get away with “different than” if a full clause follows: “Your pashmina shawl looks different than it used to since the cat slept on it.”</u>
3 (Upholds the rule, but...)	Butterfield (2015)	SPLIT INFINITIVE	7 Recommendation. , The ban on the split infinitive, though relatively recent in the broader context of the history of English, <u>has sufficient weight of opinion against it to recommend avoidance when possible</u> , and especially when it is stylistically awkward. <u>But it is neither a major error nor a grammatical blunder, and it is acceptable and at times necessary when considerations of rhythm and clarity call for it.</u>
4 (Upholds the rule)	Garner (2016)	LAY/LIE	To use <i>lay</i> without a direct object, in the sense of <i>lie</i> , is nonstandard <I want to lay down> <he was laying in the sun>. But this error is very common in speech—from the illiterate to the highly educated. In fact, some commentators believe that people make this mistake more often than any other in the English language. Others claim that it’s no longer a mistake—or even that it never was. <u>But make no mistake: using these verbs correctly is a mark of refinement.</u>

The second research question in the current study involved carrying out a series of corpus-based analyses to determine the degree to which formal edited writing (news) and informal unedited writing (personal blogs) differ in the ways they conform to the prescriptive

usage advice given for the eight usage problems identified in Section 4.1.1. I used two corpora to carry out this study, both of which I describe in the next section.

4.2 Conducting the Corpus Analysis (RQ2)

4.2.1 COCA–N

The first corpus used in this study is the news subcorpus of the Corpus of Contemporary American English (Davies, 2008). COCA is a 570-million-word monitor corpus of English that is balanced across multiple registers and freely available online at english-corpora.org. For this study, I used the full-text version of the corpus which contained 85,987,630 words in news articles published between 1990 and 2012. The corpus was tagged with the CLAWS tagger (Garside & Smith, 1997) and was available in a three-column, vertical format with each word and its part-of-speech tag(s) on a line. I wrote a script in the Python programming language to combine the words with their tags and divide the text with approximately one sentence per line. Achieving perfect accuracy in organizing the text this way was not possible because a percentage of the text in the corpus was removed by Mark Davies, the compiler, for copyright reasons. Therefore, splitting the texts into complete sentences was not possible because some portions of some of the sentences have been removed. However, I was able to approximate a one-sentence-per-line structure similar to the structure of the data in the corpus of personal blog posts that I compiled (see Section 4.2.2.3). This corpus of news writing was used to represent formal written English in the current study, as discussed in Chapter 3.

4.2.2 Corpus of Personal Blog Posts (CPBP)

The second corpus—the corpus of personal blog posts or CPBP—was compiled specifically to carry out the current study. Other corpora of blog posts have been compiled

(e.g., the Birmingham Blog Corpus and the Corpus of Global Web-Based English, 60% of which is composed of informal blogs (Davies, n.d.)); however, these corpora seem to prioritize size over collection methods guided by a careful situational analysis as is described here. Therefore, in order to ensure that the blogs included in the corpus are all personal blogs (as opposed to corporate or other blogs associated with professional organizations) and that they were written by American authors, a new corpus of personal blogs was compiled for the purposes of this study.

To build this corpus and to ensure that it is as representative of personal blogs as possible, I conducted systematic internet searches for blogs hosted on popular blogging platforms. Following methods established in previous research (e.g., Biber & Egbert, 2018; Biber, Egbert & Davies, 2015), I used frequently occurring trigrams, or commonly occurring three-word sequences (e.g., *one of the*, *a lot of*, *out of the*, etc.), as search terms in an attempt to minimize the effect of search-engine algorithms returning personalized results. I then carried out an extensive situational analysis of the URLs returned by my searches to collect publicly available information about the blog's author, and to determine whether or not the blog was in fact a personal blog. I then collected individual posts from each personal blog domain and used Amazon's Mechanical Turk crowd-sourcing platform to determine the topic and purpose of a sample of the posts I included in the corpus. I discuss each of these steps in more detail in the following sections.

4.2.2.1 Collecting URLs. Because the goal of the current study is to compare formal and informal writing, I included only blogs in my corpus that I might reasonably assume are not professionally created or edited, and thus representative of informal writing. One way I accounted for this need was to include in the corpus only blog posts that are hosted on

popular, turnkey blogging platforms. To find the blogs I included, I performed site-restricted searches (Grieve, Asnaghi & Ruetten, 2103) on three widely used blogging platforms (wordpress.com, blogspot.com, and typepad.com) using three different search engines (google.com, bing.com, and duckduckgo.com). Two of the blogging platforms I included in this study (blogspot and wordpress) allow users to host their blogs for free. Typepad offers both free basic services as well as paid services. I did not determine whether the Typepad profiles of the blogs included in my corpus were free or paid (I am not aware of a way that end users are able to get this information); however I did check each URL included in the corpus to make sure it was not mapped to a uniquely registered domain. I did this because it is likely that users who do not pay for privately hosted website domains are also likely not paying professional writers or editors to develop the content of their blogs. In other words, it is likely that the users of free or low-cost blogging software are creating their own content and publishing it themselves. After collecting the blogs, I opened each one to be sure that it did not redirect to a site that did not include either *blogspot*, *wordpress*, or *typepad* in the URL.

Collecting a completely random sample of blog posts to include in this corpus would be ideal, as it would increase the likelihood that the blogs contained in the corpus are truly representative of personal blog writing on the web. Unfortunately, collecting a completely random sample of blog posts is not possible because the search engines I used to collect the URLs in some cases use complex algorithms to return results that they believe will be most useful to their users.¹⁰ In order to account for this limitation and therefore mitigate the influence of these algorithms, I used trigrams that frequently occur in COCA (Davies, 2011)

¹⁰ DuckDuckGo may be an exception to this statement. It does not collect any data about its users, which seems to suggest that it doesn't use algorithms to return customized search results.

as search terms (see Biber & Egbert, 2018; Biber, Egbert & Davies, 2015). The trigrams consisted largely of function words (e.g., *in middle of*, *the kind of*, *on the other*) so they contained minimal lexical meaning. Conducting the searches in this way minimized any effect of search engines' tendency to return personalized results, thereby making the results the search engine returned closer to a random sample. In addition, knowing that these trigrams are highly frequent across English registers helped to ensure that the search engines returned results from a wide variety of URLs and not only from highly specialized or niche websites. I used three different search engines to collect the data because a pilot analysis showed that different search engines return different URLs for the same search terms. In my data, I found that more than 98% of the URLs my searches returned were returned by only one search engine. That is, when conducting the searches as I describe below, the vast majority of URLs returned by each search engine were unique. Had I used only one search engine to gather URLs, the sample of blogs I collected would have been much more limited in scope.

Table 4.4 shows the trigrams I used to perform searches with each search engine on each blogging platform. I conducted the searches on August 31, 2018, and September 3, 2018. These trigrams were selected from the 100 most frequently occurring trigrams (not case-sensitive) in COCA (Davies, 2011) and randomly assigned to each cell in the table. Through these searches, I collected 3,000 URLs using each of the search engines and 3,000 URLs from each of the blogging platforms, resulting in 9,000 URLs total.

Table 4.4 Trigrams used in online searches

	Blogspot	Typepad	WordPress	TOTAL
Google	1,500 URLs (100 URLs for each of these trigrams) <ol style="list-style-type: none"> 1. I want to 2. the number of 3. I think that 4. he didn't 5. in the middle 6. don't know 7. I can't 8. the kind of 9. going to have 10. of the world 11. back to the 12. because of the 13. that it was 14. more than a 15. at the same 	1,500 URLs (100 URLs for each of these trigrams) <ol style="list-style-type: none"> 1. the middle of 2. one of the 3. as well as 4. a lot of 5. in terms of 6. not going to 7. in the united 8. the fact that 9. the end of 10. and i think 11. you can't 12. in new york 13. you don't 14. most of the 15. I couldn't 	--	3,000
Bing	800 URLs (100 URLs for each of these trigrams) <ol style="list-style-type: none"> 1. I think the 2. there is a 3. be able to 4. would have been 5. for the first 6. we don't 7. all of the 8. don't think 	700 URLs (100 URLs for each of these trigrams) <ol style="list-style-type: none"> 1. going to be 2. it was a 3. side of the 4. was going to 5. it would be 6. rest of the 7. the united states 	1,500 URLs (100 URLs for each of these trigrams) <ol style="list-style-type: none"> 1. part of the 2. end of the 3. at the time 4. to be a 5. they don't 6. is one of 7. on the other 8. of the most 9. in the first 10. at the end 11. in the world 12. in order to 13. some of the 14. the rest of 15. there was a 	3,000

Table 4.4 continued

DuckDuckGo	700 URLs (100 URLs for each of these trigrams)	800 URLs (100 URLs for each of these trigrams)	1,500 URLs (100 URLs for each of these trigrams)	3,000
	1. the first time 2. the use of 3. to go to 4. is going to 5. I have to 6. to make a 7. what do you	1. didn't know 2. to be the 3. to do it 4. this is a 5. are going to 6. didn't have 7. I don't 8. to do with	1. do you think 2. the same time 3. I didn't 4. I think it 5. in front of 6. don't want 7. you want to 8. that he was 9. a number of 10. you have to 11. to have a 12. don't have 13. out of the 14. percent of the 15. a couple of	
TOTAL	3,000	3,000	3,000	9,000

I did not use Google to collect any posts hosted on wordpress.com because I observed through piloting this collection method that Google returned many results for sites that are not blogs but that are nevertheless hosted on the WordPress platform. For instance, some of the WordPress searches I carried out using Google returned URLs to WordPress sites that redirected to another URL, often the website for a local news channel. The other search engines did not seem to perform in the same way to the same extent. They returned mostly those results that maintained “wordpress.com” in the URL, and were thus more likely to be personal blog posts than some other online register.

4.2.2.2 Determining author location and status as personal blog. The searches I conducted as described in the previous section returned URLs to blogs on the domain level (e.g., <http://ourodysssey.blogspot.com/>) as well as URLs to individual posts (e.g., <http://microbrewreviews.blogspot.com/2017/07/the-fine-art-of-sequential-credits-you.html>). In order to determine the geographic location of the authors, I identified all of the unique domain-level URLs in the 9,000 URLs I collected. Out of 9,000 total URLs, 5,732 unique

domains were identified. The next step involved opening each of these URLs to determine if the author disclosed their location somewhere in their blog or in their other publicly available online profiles. Determining the location of the authors was important because regional variation exists in written American English (Grieve, 2016). Therefore, in order to increase the representativeness of the corpus, it was necessary to be sure that it included texts from authors coming from all of the dialect regions Grieve identified. A pilot analysis revealed that author location was available in many cases, but it was not structured in a way that made automatic retrieval feasible. Therefore, I worked with 16 trained volunteers and 52 paid Mechanical Turk workers¹¹ to analyze all 5,732 URLs to determine if information about the author's location was available. To carry out this analysis, we examined a variety of sources of information, including the blogger's profile page, their publicly available social media accounts, and in some instances individual posts on the blog to find information about their location. Some bloggers shared general information about their location (e.g., their region or country), but not at the level of specificity required for this study. In order to be included in the corpus, the blogger needed to identify their location at least at the state level. Many bloggers also identified the city, but that level of detail was not necessary to be included in the corpus.

I used the information about the state that the author shared to determine their region, according to Grieve's (2016) proposed dialect boundaries, shown in Figure 4.3.

¹¹ I had two Mechanical Turk workers evaluate each URL, but I did not require the responses from both coders to match in order for the data to be usable. Having two coders served as more of a back-up. In instances in which one coder may not have found the data, the second coder might have.



Figure 4.3 Map of American dialect regions. Source: Grieve, J. W. (2016). *Regional variation in written American English*. Cambridge University Press. Used with permission.

In some cases, Grieve’s dialect boundary lines dissected individual states in a way that divided a sizable portion of the state into more than one dialect region. For instance, the eastern half of Virginia is part of the Northeast region while the western half of Virginia is part of the Southeast region. While assigning the dialect region for bloggers from these states, I used the city if was given to determine the region. For example, if a blogger specifically identified their city as El Paso, Texas, I assigned the blog to the West region. All other bloggers from Texas—including those for whom no city was identified—were assigned to the South Central region. In total, there were seven states for which that I considered the city (if available) when assigning the region. Where no city was given for these seven states, I simply assigned the blogs to the following region for each state:

- Louisiana (Southeast)
- Tennessee (South Central)
- Mississippi (Southeast)

- Virginia (Northeast)
- West Virginia (Northeast)
- Missouri (Midwest)
- Texas (South Central)

Assigning regional information this way is a suitable approach because dialect boundaries themselves are not precisely defined. In determining these boundaries, Grieve (2016) noted that the lines on the map are hand-drawn, and he cautioned against overinterpreting them. Specifically, he said that “[t]he lines represent the estimated spines of areas of transition between the five dialect regions, as opposed to definite or absolute borders” (p. 209). Thus, working with approximate assignments to different regions does not invalidate this analysis. In total, usable location information for nearly one-third ($n = 1,866$) of the 5,732 unique domains was identified. Eleven of these URLs were written by bloggers from Alaska ($n = 8$) and Hawaii ($n = 3$). Because the dialect regions of Grieve’s study included only the lower 48 states, these URLs were excluded from the analysis, resulting in 1,855 usable URLs.

One important limitation to note at this point is a latent selection bias that using this location-identification method introduced. In some cases, for URLs that appeared to be particularly good examples of informal personal blogs, more effort was spent locating the author’s location. For blogs that did not appear to represent the kind of blogs I aimed to include in this study, potentially less effort was expended in finding the author’s location. This means that there may be blogs that were excluded because no information about the authors’ location was found even though it might exist somewhere on the internet. In general, however, the efforts we made represent a responsible attempt to gather the location information of the bloggers included in the corpus.

Another limitation to address with this method is that it is impossible to know the extent to which the location information we gathered influences the authors' regional dialects. Some of the authors whose texts are included in the corpus disclosed that they have lived in many areas all across the country, yet we collected information about only one of these locations. While most of the location information we collected presumably represents the authors' current locations, this information still does not account for dialect features they may have acquired while living in a different region of the country. However, Grieve (2016, pp. 22–23) argues that length of residence should not be seen as a methodological requirement in dialectology studies. In addition, gathering this information is still valuable, as it provides much more insight into where the authors of these blogs come from or where they are currently located than other corpora of blog posts offer.

After collecting location information about the blogs, I then reviewed all 1,855 URLs to determine if the blogs could be considered personal blogs or not. The *Oxford Dictionary of the Internet* (3rd edition) defines a personal blog as “A blog that is created and maintained by one user” (“personal blog”, 2013). While highly general in scope, this definition provided a useful framework for analyzing the remaining URLs to determine whether or not they could be considered a personal blog. Some of the blogs were identified as something other than personal blogs during the location-identification task. For instance, if an analyst noted that a blog was maintained by more than one user (i.e., it had more than one contributor), they would note it in the data. For the majority of the blogs, however, it was necessary to open each URL and determine if the blog could be considered a personal blog. I personally carried out this task. Unlike the objective task of finding location information, dividing such a subjective task among multiple analysts might have resulted in inconsistencies among the

riters. In addition, looking more carefully at the data allowed me as the primary researcher to gain more familiarity with it and to iteratively determine the criteria with which I could determine the blog's status as personal or not.

In order for a blog to be considered personal, it needed to exhibit at least some of the characteristics of a personal blog (as opposed to an organizational blog or some other kind of blog). These characteristics included

- a picture of the blog owner
- a linked profile page, biographical page, or “about me” page that provided details about the author and suggested that the blog was maintained by one user
- posts with a person's name attributed to them in the metadata (the “posted by” information at the bottom of the post)
- a URL that did not redirect to a site that no longer contained *wordpress*, *blogspot*, or *typepad* in the URL
- a lack of discounting qualifiers such as evidence of having multiple contributors, posts that are poems or lyrics instead of prose, class blogs that teachers used to manage their classroom activities, and the like.

While the above items served as useful identifiers of personal blogs, not all blogs that met these criteria were considered personal blogs. In some cases, a blog could easily be identified as personal, for instance online diaries in which a writer simply detailed their thoughts about goings on easily met many of the criteria described above. Other times, the blog met some of the criteria above, but the content of the blog was clearly thematic—sometimes seemingly related to a hobby or even a profession. In these cases, I made a subjective determination as

to whether the blog seemed more tied to the person or to the organization. Figure 4.4 and Figure 4.5 show the homepages of a personal blog and an organizational blog, respectively.

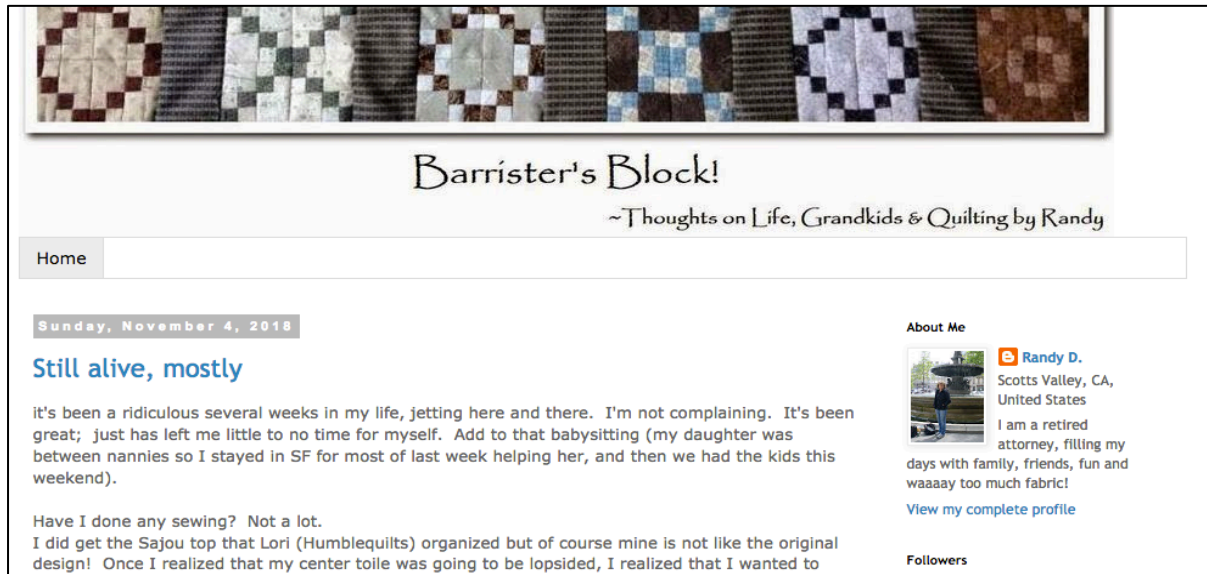


Figure 4.4 Homepage of a personal blog included in this study

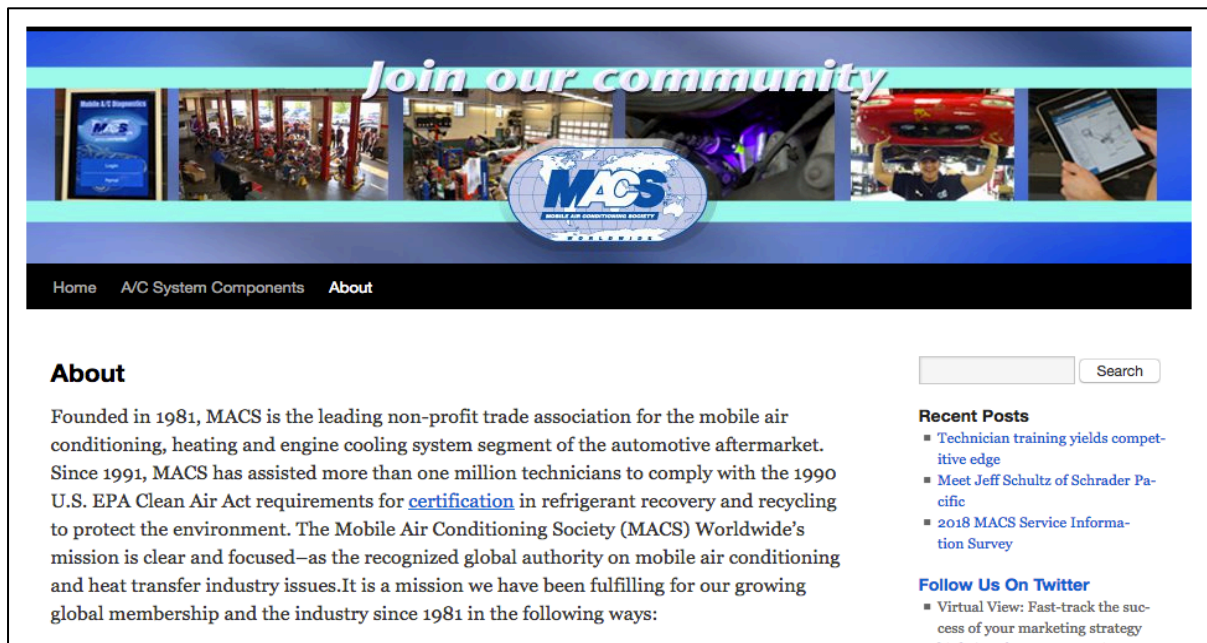


Figure 4.5 Homepage of an organizational blog excluded from this study

I did not include blogs that were clearly written as school assignments or were used as classroom management tools, as the purpose for these blogs differs from the purpose of other personal blogs.

After conducting this analysis, 1,337 usable URLs were identified, meaning that 1,337 of the original 9,000 URLs I collected were personal blogs that contained information about the location of the author. A summary of the numbers of blogs collected using this method is presented in Table 4.5.

Table 4.5 Summary of the steps taken to collect the blog URLs used in this study

Step	Description	Number
1	URLs collected from search engines on three blogging platforms. This included domain-level URLs and post-level URLs, and it included 30 URLs from the Google searches that linked to the images page for the search terms.	9,000
2	Domain-level URLs left after removing duplicate domain-level URLs (including the one remaining URL for google.com from the image search terms noted above). These are the URLs that were studied to find information about the location of the author.	5,732
3	URLs of US bloggers with at least the state disclosed which were studied to determine whether they were personal blogs or not.	1,855
4	URLs identified as personal blogs with author location disclosed. These are the URLs that are included in the corpus. 654 are from blogspot, 337 are from wordpress, and 346 are from typepad.	1,337

4.2.2.3 Scraping, cleaning, and tagging the texts. As mentioned previously, I used domain-level URLs to identify the authors' locations and to determine whether each blog was a personal blog or not. However, each text in the corpus is composed of a single individual blog *post*. That is, each text in the corpus contains one post from one author. To collect the posts from each blog domain, I scraped the posts from each blog's RSS feed if it was available. RSS feeds structure online content into XML-formatted data, which allows it to be easily displayed in RSS readers like Feedly or Inoreader. Web users can subscribe to sites' RSS feeds using an RSS reader, which allows them to easily see which of their favorite sites have updated content without having to check each individual site. For the purposes of

this research, the RSS feeds to the blogs allowed me to scrape the posts that were available in the RSS feed in a relatively simple, structured way.

RSS feeds can be accessed by URLs that are different from the URLs a person would visit to access a website. The URLs to the RSS feeds for blogs hosted on Blogspot and WordPress are standardized. To get the RSS URLs for blogs hosted on Blogspot, I simply appended `/feeds/posts/default?alt=rss` to the end of the domain level URL. To get RSS URLs for blogs hosted on WordPress, I appended `/feed/` to the end of the URL. RSS feeds for blogs hosted on Typepad demonstrated some variation, so to collect the RSS feeds for the blogs hosted on Typepad, I wrote a script that searched the source code of each Typepad blog and located an RSS URL if one was available. This script found URLs to RSS feeds for all but five blogs hosted on Typepad. I then manually checked the source code for the five Typepad URLs for which my script did not find a match and found that only one did not have an RSS feed enabled. I added the links to the RSS feeds of the other four blogs to the dataset.

In total, then, I was able to collect RSS feeds for 1,336 of the 1,337 domain-level URLs with location information available and that I identified as personal blogs. The RSS feed for each domain contained the contents of the most recent posts. Users are able to determine how many posts they want to be included in the RSS feed. The number of posts from each domain-level URL ranged from one post to 126 posts (mean = 17.96, SD = 9.77, median = 25). I then wrote a script that scraped the publication date, title, and post content of the blogs available in the RSS feed for each domain-level URL. As mentioned above, each individual blog post was written into a separate text file. I used the BeautifulSoup library to remove HTML mark-up from the post content.¹² Some RSS feeds contained only a preview

¹² Later analyses of the corpus revealed a small amount of noise that resulted from the removal of the HTML. Some of the words in the blog posts were separated only with HTML code and not spaces. In these cases, after

of the post text instead of the full post. In total, 86 domain-level URLs did not contain the post content in the expected place in the RSS feed and were therefore not included in the final corpus. The RSS feeds for 26 additional URLs were either broken, missing, or not functional for some other reason, leaving a total of 1,225 unique blogs represented in the final corpus with a total of 22,001 individual posts included (13,202,642 million words¹³). The texts were then cleaned and tagged with the CLAWS tagger using the CLAWS7 tagset (Garside & Smith, 1997) available in the ClawsAnt program (Anthony, n.d.). The complete CLAWS7 tagset is reproduced in Appendix A.

4.2.2.4 Determining the purpose and topic of texts. After collecting the content of the 22,001 posts included in the CPBP, I selected a stratified random sample of 15% ($n = 3,631$) of the posts (including up to three posts per unique domain) to have coded for additional situational variables, including the purpose and the topic of each post. I used the Mechanical Turk (MTurk) crowd-sourcing network operated by Amazon for this analysis. Previous research has shown that the work done by MTurk workers is comparable to work done by workers from other populations (see Egbert, Biber & Davies, 2015). In the task, I provided workers with a link to the post, asked them to read it and then answer five questions related the purpose, topic, and amount of quoted material in the post. I paid workers \$0.12 for each survey they filled out, and I had two workers rate each post.

The reason I selected a stratified random sample of posts to code rather than having the full corpus coded was primarily due to cost constraints. In light of ethical considerations

the HTML was removed, the words on both sides of the code were run together instead of being separated by a space. This caused the tagger to mistag these words. While future versions of the corpus would account for this issue, it appeared to have occurred infrequently and likely did not substantially (if at all) impact the analysis presented here.

¹³ Calculated with AntConc (Anthony, 2104) with the “Hide tags” setting activated.

related to paying MTurk workers (see Semuels, 2018), it was important to me to pay what I felt was a fair wage for the work that I was requesting. Therefore, rather than trying to code most or all of my corpus for much less money, I prioritized paying fairly for work that I requested. Because the URLs that are coded come from a stratified random sample of the full corpus, they are representative of the entire corpus and can still offer valuable insights into the situational characteristics of the corpus.

The full survey instrument appears in Figure 4.6 below. The first three questions in the survey ask about the purpose of the blog post: the extent to which its purpose is to narrate or report on events (past or present), to express opinion, or to describe or explain information. These questions were based on Biber and Egbert's (2018) call for future research to carry out this type of situational analysis that considers register as a continuous construct (p. 214). The categories themselves (narrate, describe, and express opinion) were also based on Biber and Egbert's (2018) analysis of web registers (see chs. 5–7 inclusive). The categories in my survey were not mutually exclusive, meaning that a single post could narrate past events to a great degree and describe information to a great degree. Workers were not given specific definitions of “narrate,” “describe,” and “express opinion” because part of the aim of the study was to assess end users' opinion of what it means for a text to narrate, describe, and express opinion. However, users were given examples of attributes that were characteristic for these kinds of texts. For instance, workers were told that “Posts in which the author reviews a product or service, comments on current events, offers advice, or shares faith or testimony” are characteristic of the Express Opinion category.

BLOG POST LINK: `#{BLOG_URL}`

- ☐ If the link is broken, click here then submit
- ☐ If there is absolutely no text in the post, click here then submit

1. Is the purpose of this blog post to narrate or report on events (past or present)?

*Posts in which the author relates their own experiences or someone else's, **reports** current events, or shares a short story are characteristic of this category*

☐ not at all
 ☐ slightly
 ☐ somewhat
 ☐ very much so

2. Is the purpose of this blog post to express opinion?

*Posts in which the author reviews a product or service, **comments** on current events, offers advice, or shares faith or testimony are characteristic of this category*

☐ not at all
 ☐ slightly
 ☐ somewhat
 ☐ very much so

3. Is the purpose of this blog post to describe or explain information?

Posts in which the author describes how to accomplish a task or explains information about a particular topic are characteristic of this category

☐ not at all
 ☐ slightly
 ☐ somewhat
 ☐ very much so

4. What is the primary topic of this blog post?

☐ Current Events *e.g., news, politics, law, crime*
☐ Education *e.g., early childhood, elementary, secondary, university*
☐ Health *e.g., medicine, food, drink, exercise*
☐ Hobbies *e.g., crafts, travel, sports, quilting, photography, firearms, collecting, cars, toys*
☐ Humanities *e.g., language, writing, literature, history, philosophy, religion, art*
☐ Lifestyle *e.g., fashion, beauty, home & garden*
☐ Money *e.g., business, economics, finance, taxes*
☐ Popular Culture *e.g., books, TV, movies, music, games*
☐ Relationships *e.g., family, parenting, kids, dating, romance, pets*
☐ Science & Technology *e.g., math, engineering, computers*
☐ Self *e.g., self-reflection, self-awareness, self-improvement, self help*
☐ Other (please specify in one or two words)

5. About how much of the text is quoted from other sources?

☐ less than 10%
 ☐ between 10% and 25%
 ☐ between 25% and 50%
 ☐ more than 50%

Figure 4.6 Survey instrument used by Mechanical Turk workers to determine the purpose, topic, and amount of quoted material in a sample of blog posts

I also asked workers to select the general topic of each post. My goal was to create a list of topics that was both broad enough to capture the topical variation that can be expected in blog posts, while at the same time being composed of few enough items so as to still be analyzed in a manageable way. I considered using topical indexes from the Dewey Decimal system and the Library of Congress, but found that in general, these systems were too large and yet likely not specific enough for the types of topics people might write about in their personal blogs. In piloting this study, I worked with a list of 24 topics that I had adapted from the topics used by WordPress to organize the blogs curated at discover.wordpress.com. In this pilot analysis, the level of agreement was relatively poor (53% for 1,199 unique tasks), suggesting some room for improving the framework. For the current study, I modified that same list by including topics that workers frequently added to the list during the pilot study and also by looking for ways to collapse individual topics to result in fewer categories. The final list included 11 categories with an additional “other” category that users could use if they felt that none of the options fit. Along with each of these categories, I included examples of subcategories that fit within the major category.

To help ensure quality work, I required MTurk workers who participated in my project to have a master’s qualification status, meaning they are experienced workers who have been evaluated positively by requesters in the past. Unfortunately, this did not fully prevent low-quality work, as I did encounter one worker who completed nearly half of the tasks in a way that was not reliable. I had these tasks recoded by different workers and checked their work more carefully to help ensure its quality.

The HTML for the survey form was validated using validator.w3.org and tested in four different browsers (Firefox, Chrome, Safari, and Edge). During this process, I uncovered

one bug in the form: if a worker were to select both the “broken link” and “no text” checkboxes (as seen in Figure 4.6), and then unselect one, the other elements in the form were still marked as required elements. But simply unchecking the box and checking it again solved the issue. Because the majority of the URLs were not broken or text-free, it is unlikely that this caused a substantial problem. Future iterations of this form would address this minor bug as well as implement machine-learning algorithms to detect patterns in the data returned by MTurk workers that may be problematic.

The dimensions of both corpora, as well as the dimensions of the sample analyzed by MTurk workers, is shown in Table 4.6

Table 4.6 Composition of corpora

Corpus	Number of texts	Total words	Mean text length
COCA–N	23	85,987,630	3,738,592.61
CPBP	22,001	13,202,642	600.09
CPBP sampled	3,631	2,158,360	594.43

4.2.3 Extracting Data from Both Corpora

In this section, I describe the methods and considerations that were used to extract the data from both the COCA–N and the CPBP, determine the sample size of instances to analyze, and create the datasets. The methods described in this section apply generally to all the corpus analyses in this study. In addition to these methods, each usage problem included in this study is associated with specific research questions and search methods, which I detail in their respective sections in Chapter 5. Organizing the description of the methods in this way allowed for a more efficient and organized presentation of the process. I conclude this section with a brief description of the nature of the data and the framework I developed to

classify each of the usage problems analyzed in the study based on the observations made in the corpus analysis.

4.2.3.1 Determining whether to use specific or general search terms. In order to conduct the analyses of each usage problem, data needed to be extracted from both corpora. I accomplished this with computer programs written in Python, which are explained in more detail in the sections that follow. When determining how to extract data from a corpus, there are two different considerations researchers must consider:

1. Researchers can identify certain specific patterns of use and develop search syntax using regular expressions to find those patterns, or
2. Researchers can conduct more general searches for the variations of the patterns and create a dataset to analyze by manually coding a sample of the relevant search results.

The first scenario favors precision (i.e., all or most of the instances returned by the search will be what the researcher intended to analyze) over recall (i.e., there may be many other instances of the target feature in the corpus that the researcher's specific search did not return). The second scenario favors recall over precision.

For this study, creating search patterns that were robust enough to yield excellent precision *and* recall was not feasible. As a result, I opted to use general search terms and then create a dataset to analyze (see Section 4.2.3.3), thus favoring recall over precision. I did this because I wanted to be sure that when sampling data to analyze, I was sampling from the complete set of possible options. My sampling methodology, as explained in Section 4.2.3.3, accounted for false positives.

4.2.3.2 Determining the number of instances to sample. In some corpus studies of lexical items, researchers manually analyze all instances of the target feature they retrieve from their corpus (e.g., Fernández-Pena, 2017; Gray & Cortes, 2011; Liu & Zhan, 2015). In other studies, however, researchers code only a sample of the target feature they find in their corpus. For many of the usage problems I discuss below, the number of results returned by my corpus searches were more than could feasibly be coded and analyzed due to time and budgetary constraints. Therefore, sampling from the results was necessary, which made determining the number of instances to sample an important consideration.

In their introductory textbook on corpus linguistics, Biber, Conrad, and Reppen (1998) suggest to “code a large sample” (p. 270) when conducting a variationist study in which each linguistic feature acts as the unit of observation. They later demonstrate a multivariate analysis with a coded sample of 100 instances from two different registers, possibly suggesting that a sample size of 100 is large enough to carry out corpus research. Recent studies have adopted a slightly larger sample size. For example, in his study of *many* and *a lot of* used in positive sentences, Dichtel (2016) manually coded a sample of 200 instances of each term to carry out his investigation. And in his study of collective nouns, Lakaw (2017) randomly selected 250 instances of the target feature to analyze any time his search yielded more than 500 tokens in a given timeframe.

Based on these studies, I aimed to analyze 250 instances of each alternation from each corpus for each of the usage problems. I selected to analyze 250 instances from each corpus because doing so allowed me to analyze each usage problem in adequate depth while still allowing me to maintain the breadth of studying eight usage problems (as opposed to a

smaller number). Analyzing more than this for each usage problem would have required the breadth of the study to decrease, which was not what I wanted.

4.2.3.3 Sampling the data and creating the dataset. To create the dataset for each usage problem, I imported each of the text files created by the Python programs I wrote into Excel, and I used the RAND() function to assign a random number to each KWIC¹⁴ line. I then sorted the data by register, coded status, and random number in order to prioritize the lines that came from blog posts I had coded for purpose, topic, and amount of quoted material using Mechanical Turk (see Section 4.2.2.4). In this way, the KWIC lines I analyzed constituted a near-random sample. I then selected the first 250 instances of each variant from both corpora to manually code. Because the patterns I used to collect the data were written to favor recall over precision, as noted above, my data contained many false positives. To correct for this, I ensured that I analyzed 250 actual instances of each variant from each corpus. That is, I did not include false positives in the counts for the analyses. In total, I analyzed 6,381 KWIC lines from both corpora. A summary of the analysis is presented in Table 4.7. In the sections that follow, I describe each usage problem, the methodology and the results of my analysis for each of the usage problems included in this study.

4.2.3.4 Nature of the data. Because the current study is situated in the variationist paradigm as discussed in Chapter 1—a paradigm that in many cases requires sampling equally sized portions of data to manually code—my primary method of analyzing data was to generate proportions or percentages of prescriptively correct and incorrect variants to compare. That is, I calculated the proportion of instances that were observed to follow and

¹⁴ *KWIC* is an acronym commonly used in corpus linguistics. It stands for “keyword in context.”

Table 4.7 Variants analyzed for each usage problem and the number of variants analyzed in both corpora

Usage problem	Variant(s) Analyzed	No. instances from COCA–N*	No. instances from CPBP*	Total No. instances included in analysis
LAY/LIE	<i>LAY</i> (<i>lays, laid, laid, laying</i>)	250	250	1,500
	<i>LIE</i> (<i>lie, lies, lied, lain, lying</i>)	250	250	
	<i>lay</i> (intransitive past/transitive present)	250	250	
WHO/WHOM	<i>who</i> in subject/object position	250	250	1,000
	<i>whom</i> in subject/object position	250	250	
DIFFERENT TO/THAN/FROM	<i>different than</i> followed by a clause or a phrase	250	250	500
SPLIT INFINITIVE	Infinitives with adverbs in initial, medial, or final placement	250	250	500
I FOR ME	Conjoined NPs with one observable accusative pronoun	250	205	955
	Conjoined NPs with on observable nominative pronoun	250	250	
NONE IN PLURAL CONTEXT	<i>none</i> used in reference to a singular or plural noun and a singular or plural verb	250	250	500
LESS/FEWER	<i>less</i>	250	175	925
	<i>fewer</i>	250	250	
SINGULAR THEY	Pronoun used to refer to an indefinite pronoun	250	250	500
Total		3,250	3,130	6,380

*Not including false positives

not follow the traditional rule for each usage problem. I qualify the terms *correct* and *incorrect* with the word *prescriptively* because from a linguistic standpoint it is fruitless to label comprehensible language that successfully communicates an idea as either “correct” or “incorrect.” However, because this study is concerned with prescriptive views of language, I have adopted these terms in my analysis. Based on the observed percentages of prescriptively correct and incorrect usage in both registers, I classify each rule according to the framework I developed, presented in Table 4.8. The percentages I used in the framework are adapted from the framework Biber et al. (1999) created to classify the percentage use of regular versus irregular verb forms (p. 397). Dant (2012) references the same framework in her study of prescriptions contained in the *Chicago Manual of Style*. Biber et al. do not explicitly justify the reasons that they selected the cut-off points in their framework, which may suggest that the cut-offs were arbitrarily chosen. However, even in spite of their potentially arbitrary nature, the cut-offs Biber et al. propose still provide a useful starting point to quantitatively differentiate stages of language variation. The key difference between my framework and the others just mentioned is the labels I have applied to the different ranges (e.g., “formal,” “favors formal,” etc.). These labels are used to classify different rules based on how they are followed in the different registers included in this study. For example, if the rule for a particular usage problem is observed to be followed 60% of the time in news writing and 76% of the time in personal blog writing, the usage problem would be classified as “favors informal.” This means that this particular rule is followed proportionally more often in the informal register of personal blog writing than news writing, but the difference in the proportions is not great enough to suggest that the rule is an “informal” rule—that is, one that is followed mostly in informal writing and not in formal writing.

Table 4.8 Framework developed to classify prescriptive rules

Rule classification	Proportion of instances in which the traditional rule is upheld in news writing (formal register)		Proportion of instances in which the traditional rule is upheld in blog writing (informal register)
Formal	$\geq 75\%$	and	$\leq 50\%$
Favors formal	$\geq 75\%$	and	51%–74%
Informal	$\leq 50\%$	and	$\geq 75\%$
Favors informal	51%–74%	and	$\geq 75\%$
General rule	$\geq 75\%$	and	$\geq 75\%$
Questionable status	$\leq 50\%$	and	$\leq 50\%$
Undetermined	$\leq 75\%$ OR 51%–74%	and	51%–74% OR $\leq 75\%$

Two additional examples of how the table might be used to classify usage problems may be helpful. If a rule is followed 48% of the time in news writing and 71% of the time in personal blog writing, the rule would be classified as “undetermined” because it did not reach the thresholds necessary for it to be considered formal or informal in either register. If another usage problem, however, was observed to follow the traditional rule 81% in news writing and 35% of the time in personal blog writing, the rule would be classified as “formal” because it was followed at a very high proportion in the formal register of news and was followed only at a very low proportion in the informal register of personal blog writing.

In Chapter 5 (Sections 5.1–5.8), I provide more specific information about the methods I used to extract the data for each usage problem, and I present the specific research questions related to each usage problem.

4.3 Administering the Survey to Bloggers and Journalists (RQ3)

To study the attitudes of bloggers and news writers toward the usage problems included in the study, I developed a survey based on Ebner (2017) that consisted of five

parts. Part 1 contained questions that asked participants to provide some basic information about themselves (age, regions of the United States where they have lived, etc.). Part 2 included 10 questions about participants' composing process when writing blog posts or news articles. These questions were included to assess the amount of care (Mackiewicz, 2010) or attention (Labov, 1966; 1972) participants paid to their writing and how formal they viewed their writing. Part 3 included questions about participants' experiences with English reference materials. Part 4 asked participants how they feel about the state of the English language. Part 5 asked participants to read sentences that flout the traditional rules for the usage problems investigated in this study and share whether they felt they are acceptable or not acceptable and in which contexts. The sentences included in Part 5 were adapted from the sample sentences contained in the HUGE database, many of which come from previous usage studies (e.g., Mittins, 1970) or other publications discussing prescriptivism (e.g. Crystal, 1995). The Iowa State University Institutional Review Board reviewed the survey instrument and granted it exempt status. The full survey instrument is included in Appendix B.

To find bloggers to participate in the survey, I sent an invitation to the email addresses I collected during the blog-collection phase described in Section 4.2.2.1 as well as those I collected by running a program I developed in Python that retrieved any email addresses that were included on the blog domain pages. In total, I sent the link to 695 potential participants. To find news writers to participate in the survey, I shared the survey link to two news-related listservs (NICAR-L and IRE-L), and I purchased the contact information of 485 journalists from the online database Book Your Data. I removed 15 of the contacts because they were either not in the US or worked for companies that were not news-

related organizations (e.g., Funny Or Die). I sent follow-up reminders to both groups in an effort to collect as many responses as possible. In total, I collected 82 responses to the blogger survey and 32 responses to the journalist survey. After cleaning the data by removing entries that came from the same IP addresses and entries in which the participant did not respond to any questions beyond Part 1 (demographic information), I ended up with responses from 77 bloggers and 26 journalists.

The breakdown of participants who shared their gender and age is listed in Table 4.9 below. Note the surprisingly high median age for bloggers.

Table 4.9 Age ranges and genders of survey participants

Age range	Bloggers		Journalists	
	Female	Male	Female	Male
18–29	1	0	4	3
30–39	4	5	5	1
40–49	6	3	2	0
50–59	6	10	2	2
60–69	11	21	1	3
70–79	1	5	0	1
80–89	1	1	0	0
90–99	0	1	0	0
Total	30	46	14	10
Average	54.77	59.72	38.36	49.1
Median	58	63	35.5	54.5

Respondents came from all five US dialect regions identified by Grieve (2016) as shown in Table 4.10. This is important because when data from the survey responses are considered in aggregate, it controls for variation due to regional factors.

Table 4.10 Geographic locations of survey participants

Region	Bloggers	Journalists
Midwest	18	6
Northeast	19	6
South Central	8	3
Southeast	11	5
West	21	5
Total	77	25

In the next chapter, I turn to the results of the analysis presented in accordance with the research design just presented.

CHAPTER 5. RESULTS

In this chapter, I present the analyses of the usage problems I studied. Sections 5.1–5.8 includes the individual analyses of each of the eight usage problems included in this study. Each section includes

1. a review of the advice given in the 11 usage guides I studied and the overall prescriptivism index for each usage problem (RQ1),
2. the specific methods for extracting each feature that I used in the analysis,
3. the proportion of instances each rule is followed in both registers (RQ2), and
4. a summary of the results from the survey I conducted as described in Section 4.3 (RQ3).

Section 5.9 presents a discussion of the results, including a comparison of the patterns of usage; a summary of the attitudes of usage-guide authors, bloggers, and news writers; a synthesis of the ways in which the usage patterns and the attitude profiles aligned, and a prescriptivism profile of the usage guides.

5.1 LAY/LIE

Lay and *lie* are cognate verbs that together are considered ergative, meaning they can either take an object or not (Biber et al., 1999, p. 147). Traditionally, these two verbs have been distinguished in terms of transitivity with *lay* the transitive form and *lie* the intransitive form. Both verbs follow different inflectional patterns as well, as outlined in Table 5.1.

Table 5.1 Inflectional patterns of *lay* and *lie*.

	Transitive <i>lay</i>	Intransitive <i>lie</i>
Present	lay(s)	lie(s)
Past	laid	lay
-ed participle	laid	lain
-ing participle	laying	lying

5.1.1 Review of Advice from Usage Guides

The distinction between *lay* and *lie* is difficult for many native speakers of English to remember, resulting in many instances of usage in which the rule is flouted. Garner (2016) notes that “some commentators believe that people make this mistake more often than any other in the English language” (p. 553).

The prescription for the LAY/LIE usage problem states that *lay* and its derivatives are correctly used as transitive verbs while *lie* and its derivatives are correctly used as intransitive verbs. In spite of this rule, *lay* is often used intransitively as shown in (5)

- (5) *So, wonderful Lu was kind enough to lay in the squishy moss for me
(sarahannloreth_wordpress_00010)

The troublesome nature of this rule is compounded by the fact that both the present-tense form of *lay* and the past tense form of *lie* are the same: *lay*. As a result, (6) demonstrates prescriptively accepted usage because it is clearly marked as past tense, as demonstrated by the coordinate verb *was*. On the other hand, (7) demonstrates prescriptively incorrect usage because it requires the base form of the verb in the *to*- complement clause. Because there is no direct object in the sentence, the prescriptively correct verb should be *lie*.

- (6) It was kind of fun to watch it as I lay down on the sofa for a nap
(comfortinaninstant_typepad_00003)
- (7) *One little chick liked to lay on his mother’s back
(joanne-young_blogspot_00008)

In spite of the tricky nature of this rule, many commentators on usage insist that their readers should learn the rule and uphold it. Garner (2016), Fogarty (2008), Trask (2006), Batko (2004), and O’Conner (2009) each take highly prescriptive approaches to this rule (as

shown in Appendix C), recommending that readers follow the rule without making any qualifications for different situations. Fogarty calls the distinction between *lay* and *lie* a “hard-and-fast [rule]” (p. 49) and Garner, while acknowledging that misusing *lay* for *lie* is “common” and that some “claim that it’s no longer a mistake” (p. 553), still encourages users to take the rule seriously because “using verbs correctly is a mark of refinement” (p. 553). Though he notes *lay* misused for *lie* as a Stage 4 change, meaning that the form is “universal but is opposed on cogent grounds by a few linguistic stalwarts” (p. xxxi), of whom Garner seems to be one.

Peters (2004) acknowledges that the traditional rule is commonly broken “in casual talk in all English-speaking countries” though “the standard forms **lie/lay** are still expected in the written medium” (p. 322, boldface in original). *Merriam-Webster’s Dictionary of English Usage* (1994) (*MWDEU*) offers the most lenient advice for this usage rule of any of the guides included in this study, suggesting that “the *lay-lie* shibboleth seems to be changing its status” (p. 587). *MWDEU* then cites Bolinger in offering advice for what users should do when confronted with this issue: “Many people use *lay* for *lie*, but certain others will judge you uncultured if you do. Decide for yourself what is best for you” (p. 587).

All 11 of the usage guides I included in this study contain entries related to the LAY/LIE usage problem. Considered together, the prescriptivism index for this usage problem is 3.41 on a scale of 4, indicating a high level of prescriptivism among usage-guide authors. In fact, of all the usage problems included in this study, only the I FOR ME usage problem (see Section 5.5) is treated more prescriptively among the usage-guide authors.

5.1.2 Research Questions and Method of Extracting Data from the Corpora

To carry out the analysis for this usage problem, I investigated the following research questions:

- RQ1_{lay/lie} What proportion of instances of *lay* and its derivatives (*lays/laid/laid/laying*) are used transitively and intransitively in news and personal blog writing?
- RQ2_{lay/lie} What proportion of instances of *lie* and its derivatives (*lies/lay/lain/lying*) are used transitively and intransitively in news and personal blog writing?
- RQ3_{lay/lie} To what extent do the proportions between these two registers differ?

To identify and extract the target features from the corpora, I grouped the possible instances of *lay* and *lie* and their derivatives into three groups as shown in Table 5.2. I wrote scripts in Python that used regular expressions to find the instances of each term in each corpus. These regular expressions are also shown in Table 5.2.

Table 5.2 Search terms used to extract data from both corpora

Groups	Terms	Regular Expressions	Examples
Group 1: Transitive <i>lay</i>	<i>lays</i> (3rd person present) <i>laid</i> (past AND perfect) <i>layed</i> ¹⁵ (nonstandard past) <i>laying</i> (progressive)	<code>r"(\bays \blaid \blayed \blaying)_V\w+[@%]?"</code>	Not <u>laying</u> again the <u>foundation</u> of repentance from acts that lead to death (jesusgiveslife_blogspot_00019)
Group 2: Intransitive <i>lie</i>	<i>lie</i> (present) <i>lies</i> (3rd person present) <i>lied</i> (nonstandard past) <i>lain</i> (perfect) <i>lying</i> (progressive)	<code>r"(\blie \blies \blied \blain \blying)_V\w+[@%]?"</code>	Generally speaking, it <u>lies</u> between Marietta and Sandy Springs. (news_2001)
Group 3: <i>lay</i>	<i>lay</i> (transitive present AND intransitive past)	<code>r"\blay_V\w+[@%]?"</code>	*...you hate to have them <u>lay</u> down. (news_1995)

¹⁵ I have included searches for nonstandard spellings for past tense forms such as *layed* and *lied*, which are both attested in informal online texts represented in the GloBwE corpus (Davies, 2013).

Group 1 includes all derivatives of the lemma *LAY* except for the present-tense form *lay*. Group 2 includes the all derivatives of the lemma *LIE* except the past-tense form *lay*. Group 3 includes all instances of the form *lay* which encompasses both the present tense of *LAY* and the past tense of *LIE*. I grouped the searches in this way to facilitate the analysis. To analyze the results from Groups 1 and 2, I determined whether the verb was used transitively or intransitively based on whether there was an observable direct object in the clause and coded the instance as either prescriptively correct or incorrect depending on the form of the verb in the text. To analyze the instances from the search conducted for Group 3, however, I needed to know both the tense of the verb phrase and whether the verb was used transitively or intransitively in order to determine whether it followed the traditional rule or not. For the 500 instances I coded from this group, I coded the transitivity and the tense of the verbs. A portion of the coding instrument is shown in Figure 5.1.

	P	Q	R	S	T
preti	term	posttext	verb type	tense	result
r influence_nn1	lies_vwz_nn2	in_ii our willingness to join them in reducing military spending and in the example of c	intransitive	n/a	correct
ark_nn1 and_cc	lie_vv0	down , " he said . " She 'd still be lying in the same area when_cs we came back , som	intransitive	n/a	correct
Today it is_vbz	lying_vvg	on_ii a shelf in Schoenberg 's lab in Palo Alto . <p> Becton Dickinson , the largest need	intransitive	n/a	correct
iest would_VM	lay_VVI	his hands on the scapegoat and place all the sins of Israel onto the goat , and then tha	transitive	present	correct
re she may_VM	lay_VVI	her young , Even Your altars , O LORD of hosts , My King and my God . 4 . _.	transitive	present	correct
g , we can_VM	lay_VVI	a strong and beautiful foundation for dealing with whatever comes . Eventually when	transitive	present	correct
to Earth to_TO	lay_VVI	down His perfect life so that we can be with Him in heaven one day . Greater love hat	transitive	present	correct

Figure 5.1 Portion of the coding instrument used to analyze instances of *lay*, *lie*, and their derivatives.

In total, conducting these searches returned 15,207 results (2,748 from the blogs corpus; 12,459 from the news corpus). I then sampled 250 lines from each of the three groups from both corpora as outlined in Section 4.2.3.3 and summarized in Table 4.7. I analyzed a total of 1,852 KWIC lines, of which 352 were false positives (e.g., using *lie* in the sense of “to tell an untruth” as in (8) or *lay* in the sense of “a nonprofessional” as in (9).

(8) people lie all the time to Congress”

(spiritofaprogressive_blogspot_00018_tagged.txt)

- (9) ...and some of the old prejudice remains—doctors and lay people alike

(lymemd_blogspot_00014)

For the verbs captured in Group 3, I also noted the tense. To determine the tense, I used the part-of-speech tags assigned by the CLAWS tagger, and in some cases, I used clues in the discourse such as time adverbials (e.g., *yesterday*, *tomorrow*) and marked past- or present-tense coordinate verbs.

Instances that appeared as nonfinite relative clauses as in (10) were considered transitive.

- (10) It was a feast as we plowed through about fifty of the 60 pounds laid out for

us (docsconz_typepad_00006)

After coding the data as outlined above, I marked each line as prescriptively correct or incorrect, according to the following logic:

- Lines that included a verb from Group 1 and were coded as transitive were marked as correct (11). Lines that included a verb from Group 1 and were coded as intransitive were marked as incorrect (12).

- (11) I just laid the iron on top of the ornament for about two minutes

(ocdgs_blogspot_00017)

- (12) *We were laying on my couch watching TV and some show came on

(news_1999)

- Lines that included a verb from Group 2 and were coded as intransitive were marked as correct (13). Lines that included a verb from Group 2 and were coded as transitive were marked as incorrect (14).

- (13) It used to be that parents would lie awake at night worrying about their kids
(torahsparks_wordpress_00009)
- (14) *...and gently lies her baby boy on its soft fabric
(emerginglifeministries_blogspot_00010)
- Lines that contained the verb from Group 3 and were coded as transitive present (15) or intransitive past (16) were coded as correct. Lines that contained the verb from Group 3 and were coded as transitive past (17) or intransitive present (18) were coded as incorrect.
- (15) ...he knew where to lay his wreath (news_2006)
- (16) In one chapter, a child lay dying (barryshymns_blogspot_00019)
- (17) *...when I was sitting on the couch in my apartment and lay his head across my lap so I could stroke his snout, I knew I'd adopt him.
(petermity_typepad_00008)
- (18) *Your scallops will not lay flat if they don't have this space between them
(myretiredlifeontheprairie_typepad_00001)

5.1.3 Results of Corpus Analysis

The results of the overall analysis for both registers is shown in Figure 5.2. Both registers followed the *lay/lie* rule a high proportion of the time, though the rule was followed proportionally more in news writing (96.80%) than in blog writing (85.87%).

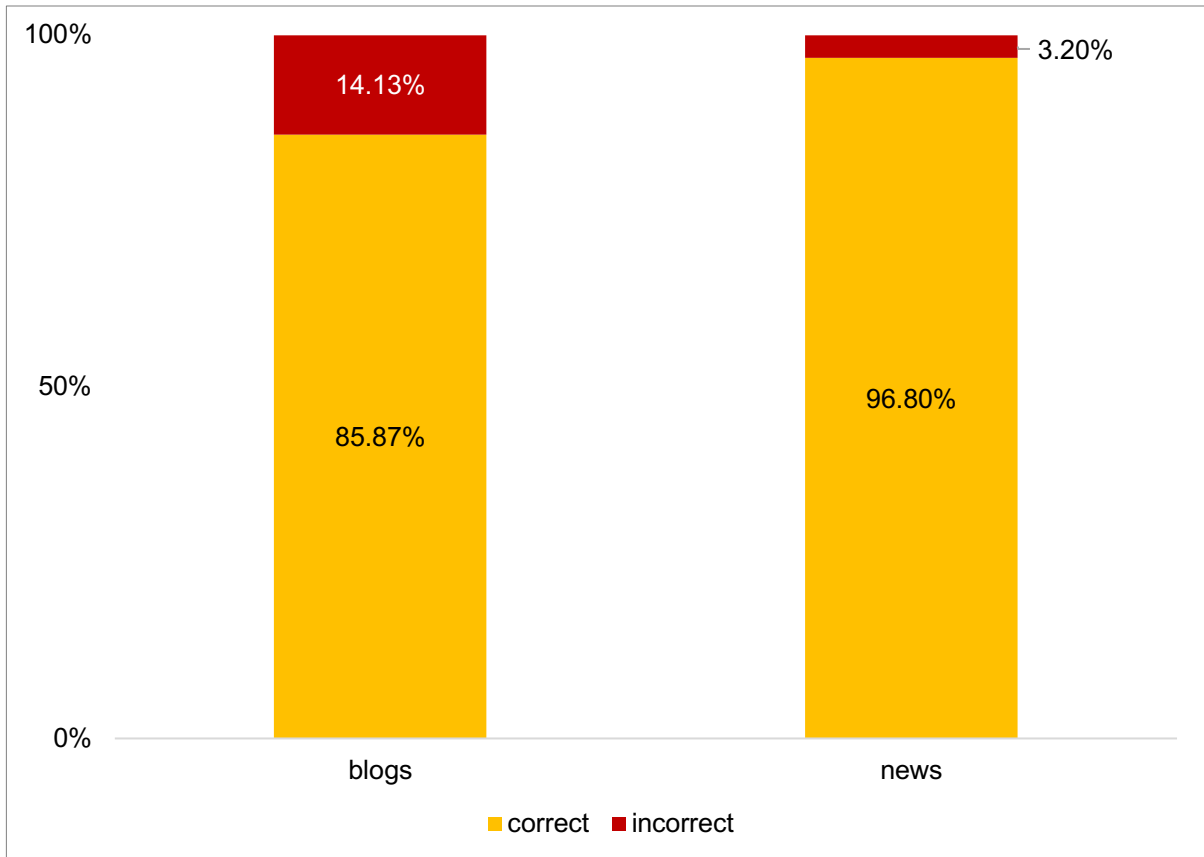


Figure 5.2 Proportion of correctly and incorrectly used instances of *lay*, *lie* and their derivatives in blog and news writing

As noted above, *lay* is the only form that is used in the inflectional patterns of both *LAY* and *LIE*: as the present-tense transitive form and the past-tense intransitive form. Because of this, one might suspect that the highest number of prescriptive errors would involve this form of the verb. However, the other forms of *LAY* were just as problematic for the writers in both corpora as shown in Figure 5.3. The rate of prescriptively incorrect usage was near 13% in both Group 1 and Group 3, indicating that writers from both groups flouted the rule at almost the same rate with derivatives of *LAY* (e.g., *lays*, *laying*, *laid*) as they did with the word *lay* itself.

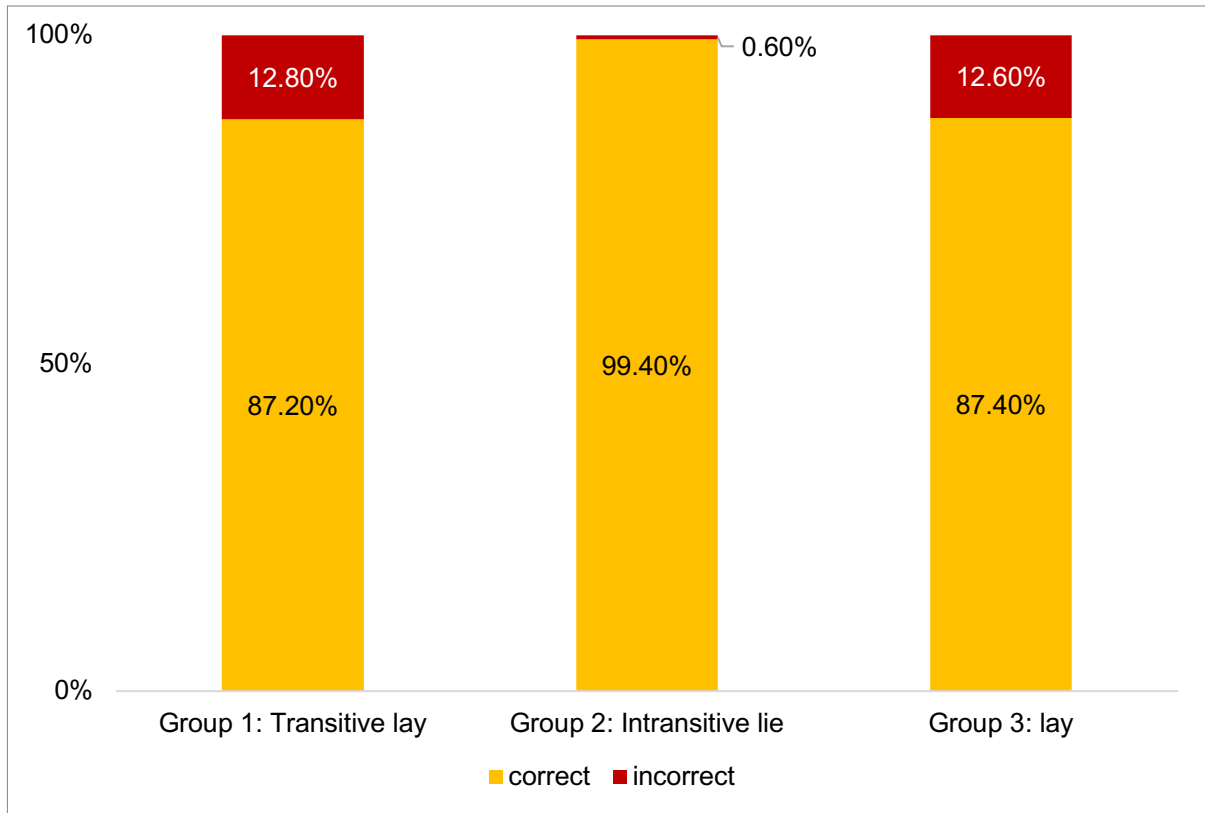


Figure 5.3 Proportion of errors in each verb group

Many of the instances from the news corpus that were marked as incorrect appeared in quoted text.¹⁶ An example of this kind of quoted material from news writing is shown in (19).

- (19) **“I wrote him off as dead at mile 40,” he said. “He pulled in there after many, many hours, and far behind his projected pace, and laid down on the gravel parking lot and just laid there.”* (news_1999)

The *Associated Press Stylebook*, the standard style guide for news writing, tells journalists that they should “[n]ever alter quotations even to correct minor grammatical errors or word usage” (quotations in the news, n.d.). This means that when an interviewee confuses

¹⁶ Quoted language is an important feature of news writing and was left in the news corpus for this study. However, leaving this quoted material in the corpus also posed a limitation for the current study because news writers do not produce the language of quoted material themselves. Future studies in which quoted material is removed from news writing could control for this variable. I discuss this topic further in Section 6.1.3.

the traditional *lay/lie* distinction, it is unlikely that a journalist or copy editor will correct the language so that it conforms to a particular style. With this guidance from the *AP Stylebook* in mind and considering the relatively high number of uses of *lay/lie* that appear in quoted material—and considering that, as noted before, confusing this rule is extremely common in English—it is surprising that the rule is observed as often as it is across registers.

5.1.4 Summary of Survey Results

To better understand attitudes of bloggers and journalists toward this usage problem, survey participants were asked whether Sentence (20) was acceptable or unacceptable, and if acceptable, in which contexts (formal writing, formal speaking, formal online communication, informal writing, informal speaking, informal online communication):

(20) *The tools were just laying there.

This sentence flouts the traditional rule because it uses the *-ing* participle *laying* as an intransitive verb.

Of the bloggers who responded to this question in the survey, 54.17% ($n = 39$) said that this usage was unacceptable regardless of context. More than one-third (34.72%, $n = 25$) felt that it was appropriate in informal written and spoken contexts. Only 11.11% ($n = 8$) felt that it was appropriate in formal writing.

The patterns among journalists were similar, though journalists were slightly more lenient in their views of this rule. Just less than half of journalists (45.45%, $n = 10$) found the sentence to be unacceptable regardless of context. More than one-third (36.36%, $n = 8$) felt that the usage was acceptable in informal speaking and writing, and 18.18% ($n = 4$) felt that it was appropriate in formal settings. In many cases, respondents who marked that a particular usage was acceptable in formal settings also marked it as acceptable in informal

settings. It may be safe to assume that even when respondents marked that a proscribed usage was acceptable only in formal settings, they might still find these examples acceptable in informal settings as well. These results are represented visually in Figure 5.4.

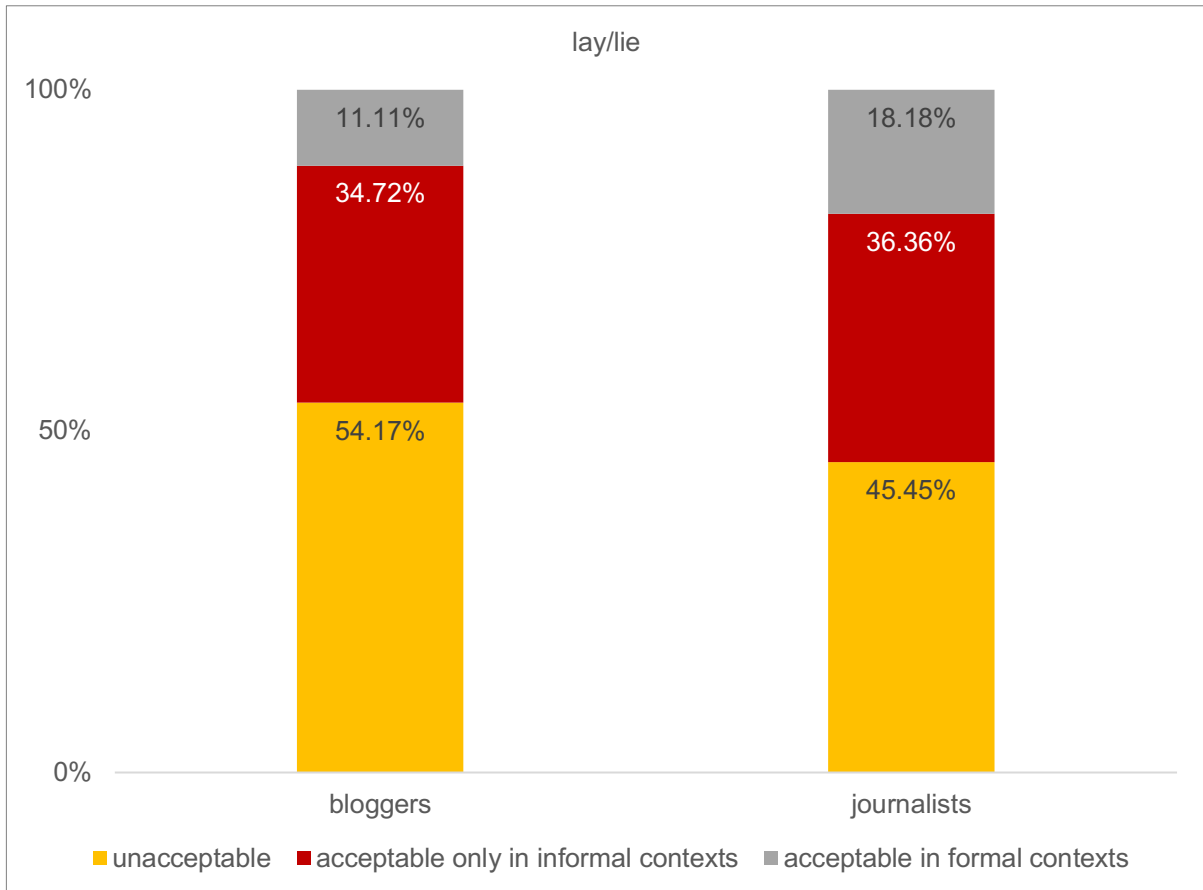


Figure 5.4 Proportion of bloggers and journalists who found the misuse of LAY/LIE unacceptable, acceptable only in informal contexts, or acceptable in formal contexts

While some respondents took a highly prescriptive tone in their comments, several participants from both groups recognized that the issue in the sentence had to do with the *lay/lie* distinction but indicated that they found the rule hard to remember. A comment by one of the journalists echoed the sentiments of the bloggers: “C’mon, nobody remembers how to conjugate lay vs. lie in the past tense.” Another journalist admitted that they find the rule confusing and that the sentence in the survey seemed correct, though “clearly not formally correct.”

These responses from survey participants reflect the advice given in the usage guides. The relatively high proportion of people who feel that flouting the rule is unacceptable could be influenced by the fact that many current usage guides recommend that readers learn and follow the rule.

5.2 WHO/WHOM

Who and *whom* are both pronouns that are distinguished by case: *who* is used as a nominative pronoun and *whom* is used as an accusative pronoun. The prescription for the WHO/WHOM usage problem requires this distinction to be observed. In other words, *who* is correctly used as a subject pronoun and *whom* is correctly used as an object pronoun according to the traditional rule. Therefore, according to this prescription, Sentence (21) would be considered incorrect because it uses the nominative form *who* as a relative pronoun when the gap in the relative clause is accusative.

(21) *The person who I met at the store was very kind.

5.2.1 Level of Prescriptivism in Usage Guides

Current views on this rule among usage-guide authors range from highly prescriptive to relatively lenient. No guide of those included in this study took a completely lenient stance with this usage problem. That is, no guide suggested that this is a rule that writers can break in any context. Many of the guides noted that the rule is often not followed, but that in formal situations, it should be maintained. For example, the *American Heritage Guide to Contemporary Usage and Style* (2005) notes that “[t]oday, the rules are well established as a part of formal Standard English. Nonetheless, *whom* is uncommon in speech and informal writing because of its inherently formal tone. When formality is not required, *who* generally replaces *whom*” (p. 504).

Others, however, take a more prescriptive tone with no concessions made for informal contexts. Strunk and White (2009) simply describe the rule with no mention of contexts in which it might acceptably be broken: “The personal pronouns, as well as the pronoun *who*, change form as they function as subject or object” (p. 11). Similarly, Fogarty (2008) discusses the rule in terms of right and wrong: “I want you to actually understand the right way to use these words...Knowing which word to choose also requires you to know the difference between subject and object because you use *who* when you are referring to the subject of a clause and *whom* when you are referring to the object of a clause” (p. 50).

While *who* has been slowly overtaking *whom* in many of its grammatical contexts, research has found that in certain grammatical contexts, *whom* is still more common than *who*. For instance, Walsh and Walsh (1989) found that when participants were asked to complete a sentence with either *who* or *whom*, in situations where the blank was the object of a preposition and the preposition is fronted, 100% of students filled it with *whom*. On the other hand, Trask (2006) calls *whom* “all but dead in English, even in formal written English” though he concedes that “the only case in which *whom* is normal and proper” (p. 285) is when it is used as the object of a preposition with the preposition fronted as in (22)

- (22) ...and the doctor for whom Lichtenstein worked was a guest speaker
(informationfornurses_blogspot_00006)

Sometimes, this can result in a doubling of the preposition as in (23)

- (23) *...chatting with a young man of whom he is not introduced to.
(oswaldsmother_blogspot_00018)

Garner (2016) takes a lenient approach in allowing *who* to replace *whom*, acknowledging that in some contexts *whom* is “stilted” and that writers “who don’t know

how to use [it] should abstain in questionable contexts” (p. 476). He is less permissive, however, in allowing *whom* to replace *who*, saying that “*whom* shouldn’t be used as the subject of any finite verb” (p. 965).

All 11 of the usage guides included in this study address the WHO/WHOM usage problem. Considered together, the prescriptivism index for this usage problem is 2.64, indicating an overall middle-of-the-road evaluation of this rule.

5.2.2 Research Questions and Method of Extracting Data from the Corpora

To determine how the WHO/WHOM usage problem is observed in actual samples of formal and informal writing, I investigated the following research questions.

- RQ1_{who/whom} What proportion of instances of *who* are used in subject position versus object position in personal blogs and news writing?
- RQ2_{who/whom} What proportion of instances of *whom* are used in subject position versus object position in personal blogs and news writing?
- RQ3_{who/whom} To what extent do the proportions between these two registers differ?
- RQ4_{who/whom} Is the rule followed more or less frequently in certain grammatical patterns?

To identify the target features, I wrote scripts in Python to extract data from the corpora using the regular expressions shown in Table 5.3 as search terms.

Table 5.3 Search terms used to extract data from both corpora

Patterns	Regular Expressions	Examples
who	<code>r"\bwho_\w+[@%]?"</code>	And <u>who's</u> going to be speaking at Advertising Week? (adcontrarian_blogspot_00010)
whom	<code>r"\bwhom_\w+[@%]?"</code>	...for Mayor Wellington Webb, <u>whom</u> Tate is careful not to criticize. (news_2003)

I then coded 250 examples of each of the two variants from each corpus, resulting in 1,000 instances coded. I used Walsh and Walsh's (1989) framework (reproduced in Appendix D) to determine the grammatical position of each instance of each variant and determined whether the variant followed the prescriptive rule or not. I observed the following grammatical positions in my data, summarized in Table 5.4:

Table 5.4 Examples of *who* and *whom* from each of the grammatical positions in which they were observed

Grammatical Position	Example
Direct object of main clause	He probably has no idea <u>whom</u> to call. (news_2007)
Direct object of relative clause	I chat with those <u>whom</u> I've never meet (2nd-cup-of-coffee_blogspot_00022)
Direct object of <i>wh</i> - complement clause	I want the world to know <u>who</u> they are praising (malikaspeaks_wordpress_00001)
Object of preposition in a main clause with the preposition fronted	They have four children, one of <u>whom</u> , Sam W. Maynes, is among the seven lawyers in his father's law firm (news_1997)
Object of a preposition in a relative clause with the preposition fronted	I recently had teachers with <u>whom</u> I work explore PYP expectations (making-teaching-visible_blogspot_00010)
Object of a preposition in a relative clause with the preposition stranded	...that no longer believes itself accountable to the people, <u>whom</u> Codevilla referred to as the "Country Class" (news_2010)
Subject of main clause	<u>Who</u> made the grasshopper? (octobia_wordpress_00001)
Subject of relative clause	These were beleaguered artisans <u>who</u> had stood up to tyranny and had prevailed (news_1992)

Table 5.4 continued

Subject of tensed clause embedded in relative clause	*How could the patriarch, <u>whom</u> I'd been told my whole life was this amazing man be willing to take on such an act? (emintheark_wordpress_00002)
Subject of <i>wh</i> - complement clause	They'll fight over <u>who</u> 's going to be top dog (news_2012)
Subject complement of main clause	In this age of global upheaval, <u>who</u> is the United States to be telling anyone what they can or cannot eat? (news_2006)
Subject Complement of <i>wh</i> - complement clause	Oh, this is <u>who</u> I am (rabbisteinlauf_blogspot_00010)

After coding the data as outlined above, I marked each KWIC line as correct or incorrect according to the following logic:

- Lines that included the nominative form *who* used in some subject position (e.g., subject of a main clause, subject of a relative clause, etc.) were coded as correct. Lines that included the nominative form *who* used in some object position (e.g., the object of a main clause, the object of a preposition, etc.) were coded as incorrect.
- Lines that included the accusative form *whom* used in some object position were marked as correct. Lines that included the accusative form *whom* used in some subject position were marked as incorrect.

5.2.3 Results of Corpus Analysis

The results of the overall analysis are shown in Figure 5.5. As the figure shows, the vast majority of uses of *who* and *whom* in both registers was consistent with the traditional rule—that is, *who* was used as a subject pronoun and *whom* was used as an object pronoun in the majority of cases.

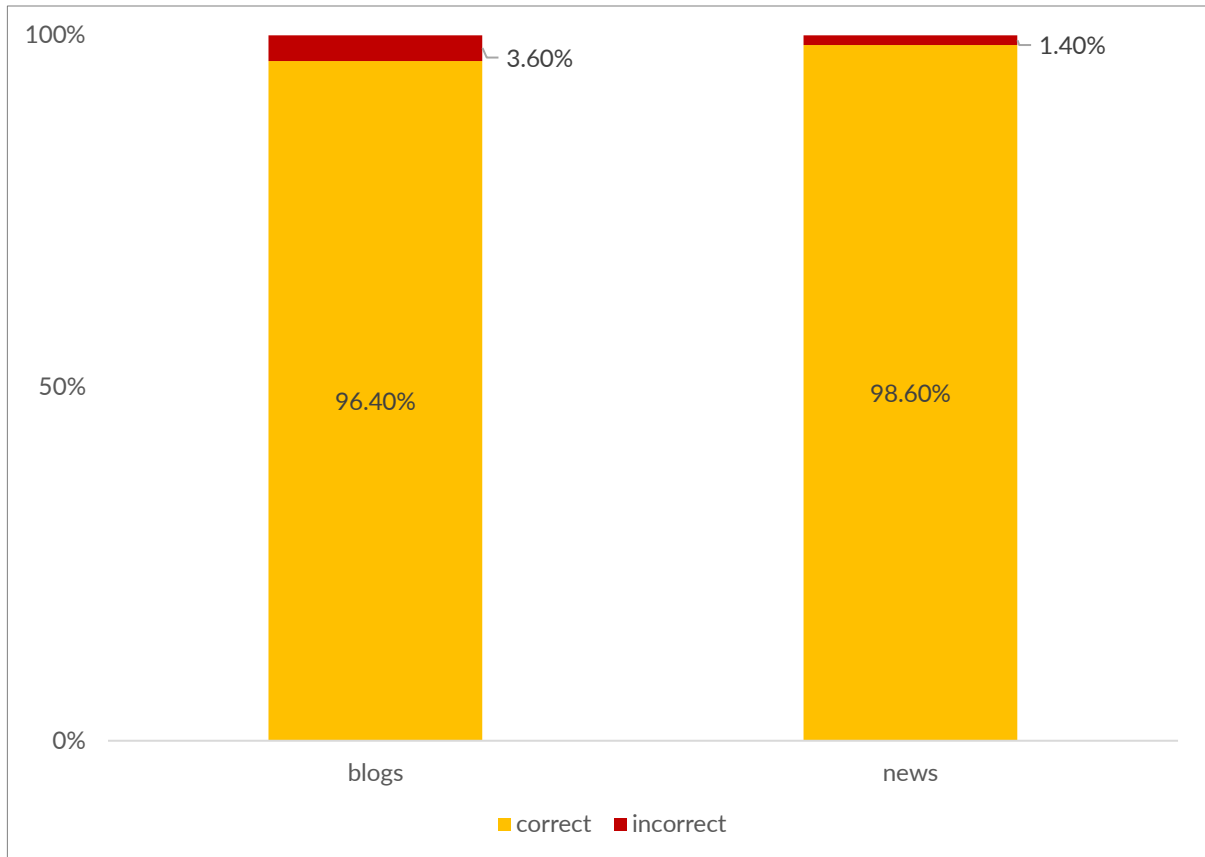


Figure 5.5 Proportion of correctly and incorrectly used instances of *who* and *whom* in blog writing and news writing

The grammatical position in which the highest proportion of nonstandard usages occurred across registers was when *whom* appeared as the subject of a tensed clause embedded in a relative clause. Sentence (24) shows an example of this kind of sentence.

- (24) *Cassio (Michael Laurence), Othello's lieutenant whom Iago strongly suggests is having an affair with Desdemona, are both somewhere here.
(wwwbillblog_blogspot_00020)

Sentence (24) is likely an example in which a writer overgeneralized a rule in an effort to be prescriptively correct. Instances such as this result in what are some usage guides call hypercorrections (see e.g., Garner, 2016; *Merriam-Webster*, 1994). Of 11 total instances of this type of grammatical construction in the data, all but one used an accusative pronoun. The

only instance that used the prescriptively correct nominative pronoun occurred in quoted material from a newspaper article:

(25) “We all need the person who you know will never let you down...”

(news_1993)

The proportions of instances in which the traditional rule was observed and flouted in both registers are shown for each grammatical category in Table 5.5.

Table 5.5 Percent of prescriptively correct and incorrect variants by grammatical category

Grammatical Position	correct	incorrect
DO of MC	1.74%	8.00%
DO of RC	13.85%	12.00%
DO of <i>wh</i> - compl cls	0.10%	12.00%
OP MC Pfronted	0.10%	0.00%
OP MC Pstranded	0.10%	0.00%
OP RC Pfronted	31.90%	0.00%
OP RC Pstranded	1.74%	0.00%
S of MC	2.46%	16.00%
S of RC	45.74%	8.00%
S of tensed clause embedded in RC	0.10%	44.00%
S of <i>wh</i> - compl cls	0.51%	0.00%
SC of MC	0.21%	0.00%
SC of <i>wh</i> - compl cls	1.44%	0.00%
Total	100.00%	100.00%

The numbers of grammatical positions in which *who* and *whom* appeared was similar across both registers. In both registers, most instances of these pronouns occurred as subjects of relative clauses. They also frequently occurred as objects of prepositions with the preposition fronted and as direct objects of relative clauses. Though the total raw frequencies are low, there were proportionally many more instances of subject complements of *wh*-clauses and subjects of main clauses in the blogs corpus. Subjects of main clauses tended to occur in interrogatives like in (26).

(26) Who suffers when this happens?

(Everythingesl-everythingesl_blogspot_00019)

Because personal blogs are generally more conversational in nature, it is logical that these kinds of interrogative clauses would occur more frequently in this register. Figure 5.6 represents the raw frequencies of each grammatical construction in both registers.

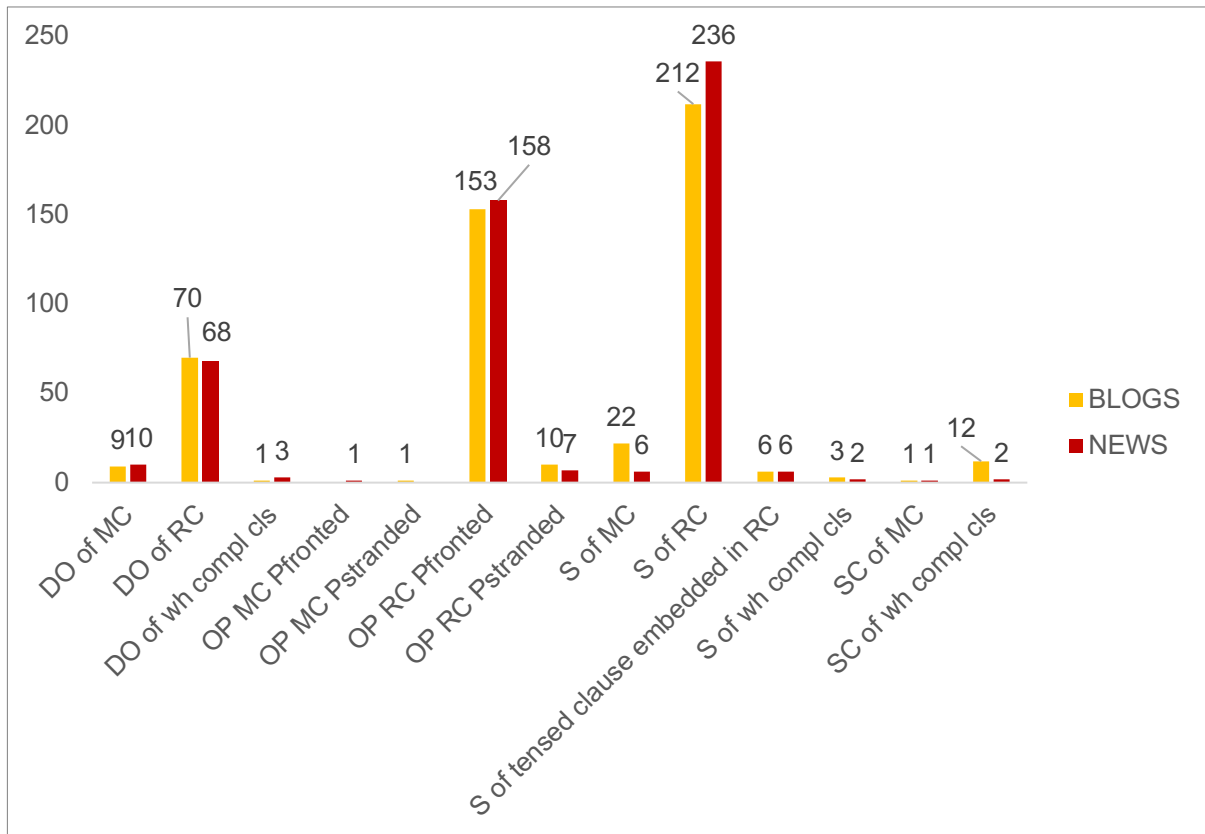


Figure 5.6 Raw frequencies of grammatical constructions in which *who* and *whom* were observed

5.2.4 Summary of Survey Results

To assess attitudes toward the WHO/WHOM usage problem, survey participants were asked whether Sentence (27) was acceptable or unacceptable, and if acceptable, in which contexts.

(27) *Who did she ask?

This sentence flouts the traditional rule because it uses the nominative *who* as the object of the main clause.

Respondents from both groups showed very similar attitudes to this sentence. Of the bloggers who responded to this question in the survey, 27.78% ($n = 20$) said that this usage was unacceptable regardless of context. Among bloggers, 45.83% ($n = 33$) found this proscribed usage acceptable in only informal contexts while 26.39% ($n = 19$) found it acceptable in some formal contexts.

Journalists' perceptions of this rule demonstrated varying degrees of acceptance.

Interestingly, the journalists' attitudes were more lenient than the bloggers with only 14.29% ($n = 3$) calling the use of *who* for *whom* in the sample sentence unacceptable in any context.

This data is visualized in Figure 5.7. This accepting attitude may be the result of journalists' general awareness of linguistic change. For instance, one journalist who noted that using *who* for *whom* as acceptable in informal speaking commented that "the rule is that 'whom' is wearing out in common usage." Indeed the use of *who* for *whom* has become so commonplace that it appears that some journalists did not recognize the fact that the sentence flouted the traditional rule as described above. One journalist who felt the question was appropriate in all informal contexts and in formal online communication commented that they thought the only issue with the sentence was that it "it does not specify 'who'."

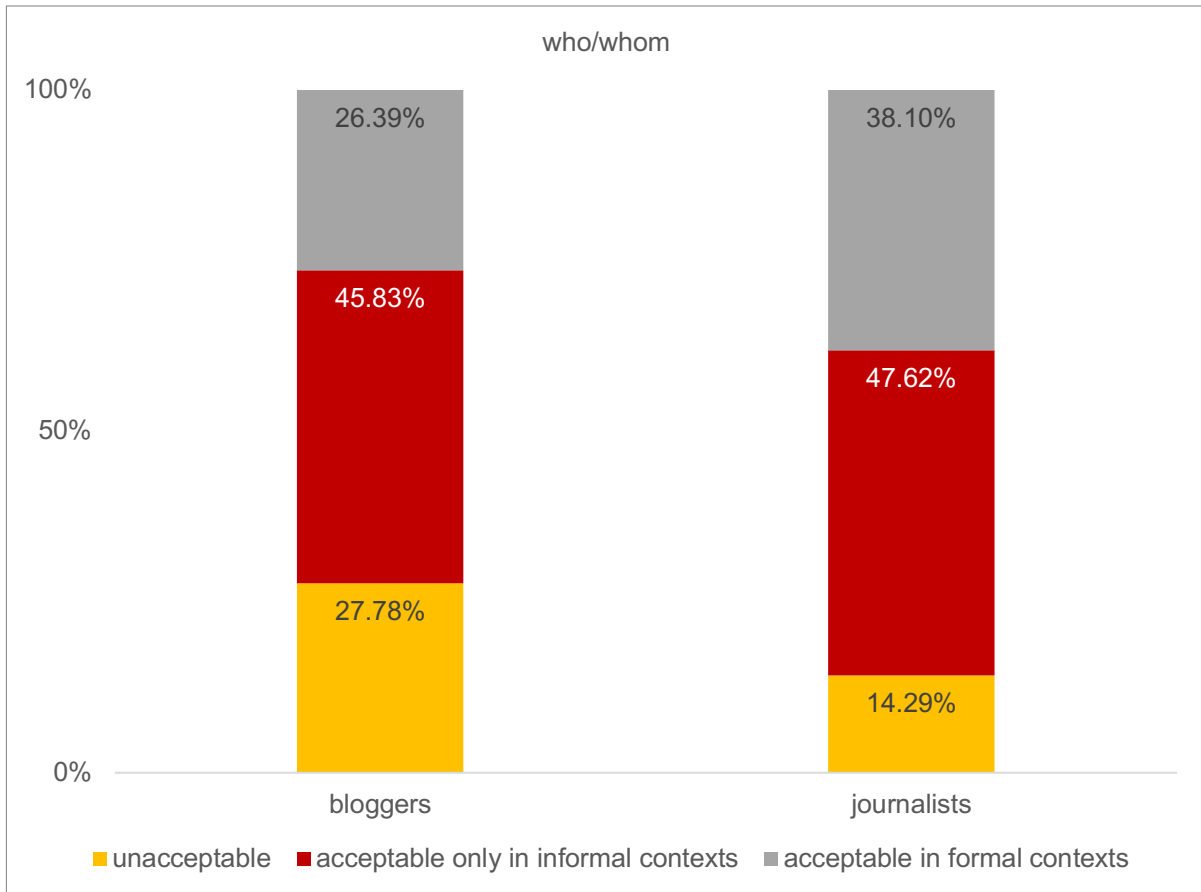


Figure 5.7 Proportion of bloggers and journalists who found the misuse of WHO/WHOM unacceptable, acceptable only in informal contexts, or acceptable in formal contexts

The general attitude of leniency and an awareness of the decline of *whom* was reflected in some of the comments of bloggers as well as journalists. One blogger, for instance, stated that “Even against former rules, ‘Whom’ now sounds affected. Furthermore, most people use ‘whom’ when it is inappropriate in order to try to sound exceptionally educated.” Some comments from journalists echoed similar sentiments. One journalist even imagined “the day when style-setters throw up their hands in surrender and say that real-world English has abandoned ‘whom...’” (ellipses in original).

These responses from survey participants deviate to some degree from the advice given in the usage guides. In general, authors of usage guides recommend that the rule should be followed in formal contexts. However, a relatively high proportion of bloggers and

particularly news writers felt that using *who* for *whom* was acceptable even in formal writing and/or speaking.

5.3 DIFFERENT TO/THAN/FROM

The prescription for the use of *different to*, *different than*, and *different from* has a history of roughly 200 years. Some, like Strunk and White (2009), argue that the preference for *different from* is supported by logic: “one thing differs *from* another, hence, *different from*” (p. 44). However, Myhill (2004a) cites Milroy and Milroy’s (1985) observation that the preference for *different from* versus *different than* “rests not on any real superiority in terms of logic, effectiveness, elegance or anything else, but on the observed usage of the ‘best people’ at that time” (p. 39). Myhill uses this observation to support his argument that prescriptivism in the United States is based on the usage of the socially powerful and is therefore socially unfair.

5.3.1 Level of Prescriptivism in Usage Guides

The issue at the center of the controversy surrounding the DIFFERENT TO/THAN/FROM usage problem is to determine which of the three possible forms are prescriptively correct: *different from*, *different to*, and *different than*. According to some usage experts, *different from* is the only acceptable alternative of the three. Payne’s (1911) guide titled *Everyday Errors in Pronunciation, Spelling, and Spoken English* states the rule succinctly: “*Different* should be followed by *from*, and not by *to* or *than*. Not ‘This work is different *than* I thought it would be,’ but different *from* what I expected” (pp. 40–41). Some guides published more recently accept *different than* as a legitimate alternation, but only when it is followed by a clause. Brians (2013), for example, notes that readers “can usually get away with ‘different than’ if a full clause follows: ‘Your pashmina shawl looks different than it used to since the

cat slept on it” (p. 83). Butterfield (2015) notes that *different to* is uncommon in American English. Because the focus of this study is on American English, I did not include instances of *different to* in my analysis.

Entries in the usage guides related to the DIFFERENT TO/THAN/FROM usage problem were analyzed in terms of the extent to which they allowed *different than* as an acceptable alternative to *different from*. Entries ranged from highly prescriptive (Fogarty, 2008; Strunk & White, 2009) to highly lenient (MWDEU, 1994). For instance, while she refrains from calling *different than* incorrect, Fogarty calls *different from* “preferred” and offers a mnemonic for remembering that it is the “best choice”: “*Different from* is preferred to *different than*. I remember this by remembering that *different* has two *f*’s and only one *t*, so the best choice between *than* and *from* is the one that starts with an *f*” (p. 22).

MWDEU (1994) on the other hand notes that *different than* is “standard in American and British usage, especially when a clause follows *than*” (p. 341). After providing an extensive summary of the history of the rule with examples, the authors summarize by saying that this particular usage problem “need have been no problem...at all, since all three expressions have been in standard use since the 16th and 17th centuries and all three continue to be in standard use” (p. 343).

Among the guides that take a more moderate approach to this usage problem, most favor following the rule and only flouting it in special circumstances. O’Connor (2009), for example, uses the qualifying word *almost* in her description of the rule, saying that “*different from* is almost always right, and *different than* is almost always wrong” (p. 96). O’Connor does not specify when *different from* might not be considered right, but she does prescribe the use of *different from* when no clause follows. Trask (2006) makes a formal/informal

distinction, noting that “formal written English requires *different from*” while “colloquial English often has *different to* or *different than*; these are familiar in speech, but they should be avoided in careful writing” (p. 88).

Taken collectively, the overall prescriptivism index for the DIFFERENT TO/THAN/FROM usage problem is 2.64, indicating that usage-guide writers recognize the rule but also allow for it to be flouted in some contexts.

5.3.2 Research Questions and Method of Extracting Data from the Corpora

Because, as noted above, *different from* is widely accepted as correct in all circumstances, I did not include it in the present analysis. Therefore, for this usage problem, I focused my investigation only on *different than* to see what proportion of instances are followed by a complete clause or not. The research questions I investigated are:

RQ1_{diff} What proportion of instances of *different than* are followed by a complete clause in formal and informal registers of writing?

RQ2_{diff} To what extent do the proportions between these two registers differ?

To identify the target feature, I used a script written in Python to extract all instances of *different than* using the regular expression shown in Table 5.6.

Table 5.6 Search term used to find instances of *different than*

Pattern	Regular Expression	Example
different than	<code>r"\bdifferent_\w+[@%]? than_\w+[@%]?"</code>	*...and they're <u>different than</u> male hearts and livers and kidneys. (news_2002)

I then created the dataset as outlined in Section 4.2.3.3 and coded the type of complement for each instance of *different than*: whether it was a clause or a phrase (e.g., noun phrase, adjective phrase, or adverb phrase.) I marked each KWIC line that included an instance of

different than followed by a clause as correct and each one that included an instance of *different than* followed by a phrase as incorrect.

5.3.3 Results of Corpus Analysis

The results of the overall analysis are shown in Figure 5.8. The prescriptive rule that *different than* is correct only when followed by a clause is flouted more often than not in both personal blog writing and news writing. In addition, the proportional breakdown of both types of complements is nearly identical in both registers with roughly one-third of the instances being coded as correct in both registers. Sentence (28) below shows an example of a sentence with a phrasal complement (marked in italics) and (29) shows an example of a sentence with a clausal complement (underlined).

- (28) *Many of the bumpers in retail stores today are different than *those seen seven years ago* (news_2011)
- (29) How would giving a puzzle to a pair of students be different than if you gave it to individuals? (buildingmathematicians_wordpress_00001)

5.3.4 Summary of Survey Results

Survey participants were asked whether the following sentence was acceptable or unacceptable, and if acceptable, in which contexts:

- (30) *Cats are very different than dogs.

The sentence flouts the traditional rule because a phrasal complement follows *different than*.

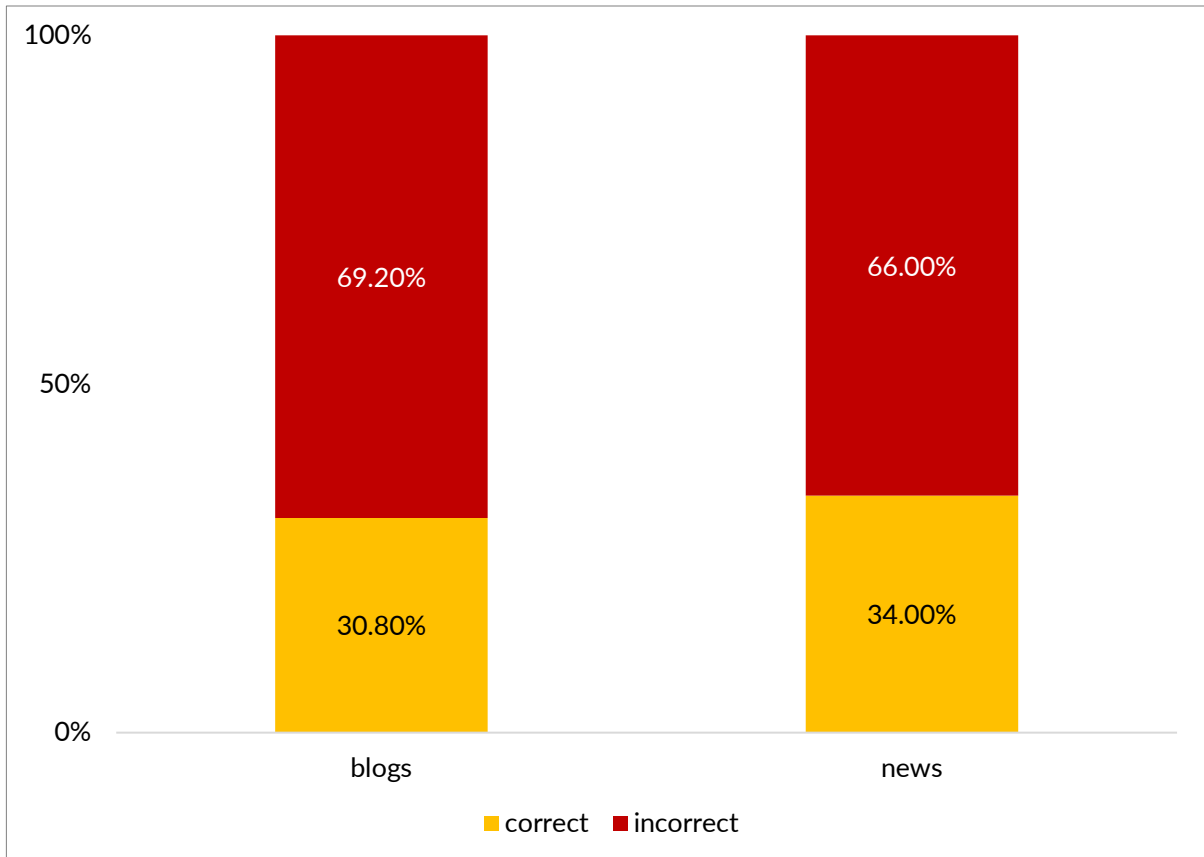


Figure 5.8 Proportion of instances of *different than* followed by a clause or a phrase in personal blog writing and news writing. Instances of *different than* followed by a clause were marked as correct; instances of *different than* followed by a phrase were marked as incorrect.

Respondents from both groups showed similar attitudes to this sentence. Of the bloggers who responded to this question in the survey, 35.71% ($n = 25$) said that this usage was unacceptable regardless of context. Among bloggers, 32.86% ($n = 23$) found this proscribed usage acceptable in only informal contexts while nearly the same proportion (31.43%, $n = 22$) found it acceptable in some formal contexts.

Journalists' perceptions of this rule resembled those of the bloggers, though their collective view was slightly more lenient than that of the bloggers. A lower proportion of journalists than bloggers, 27.27% ($n = 6$), found the use of *different than* with a phrasal complement unacceptable regardless of context. Thirty-six percent ($n = 8$) felt that the usage

was acceptable in informal speaking and writing and another 36.36% ($n = 8$) felt that it was appropriate in formal settings. These results are summarized in Figure 5.9.

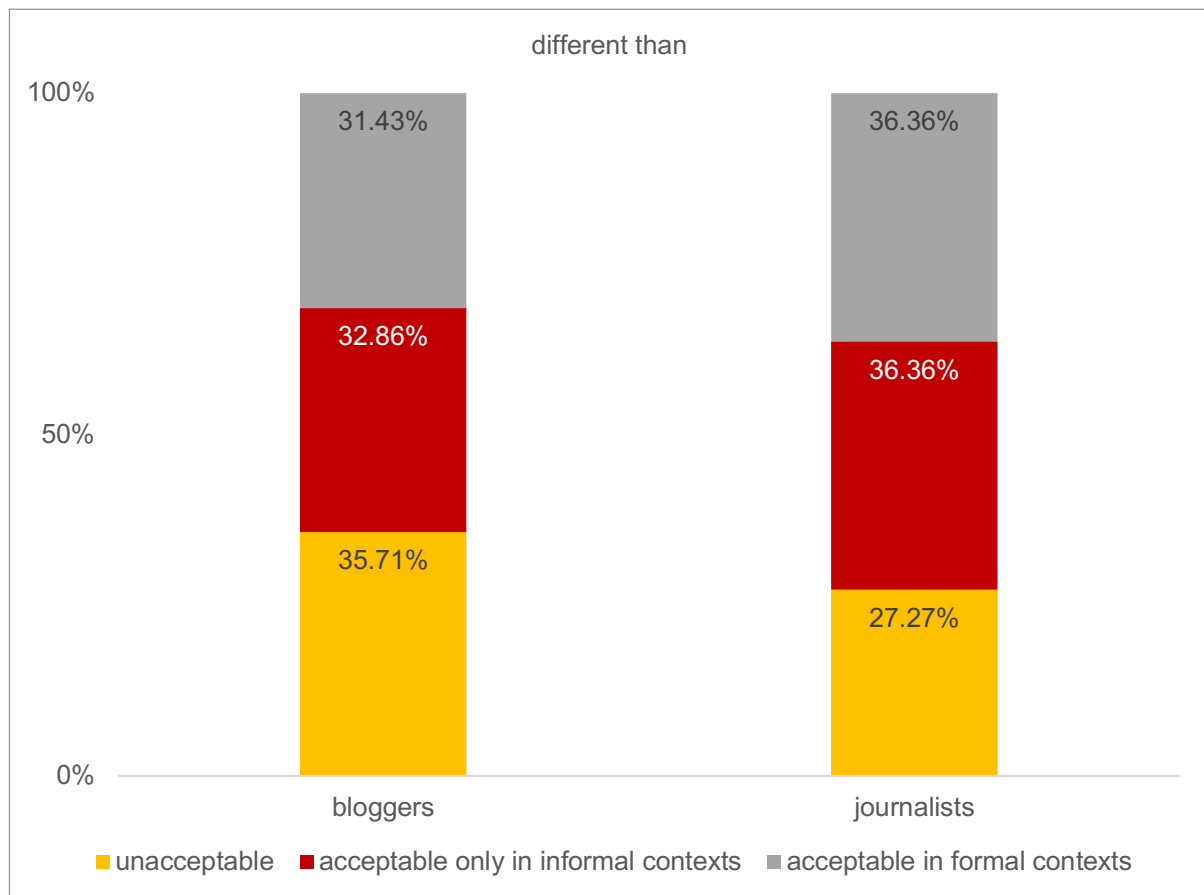


Figure 5.9 Proportion of bloggers and journalists who found the misuse of *different than* unacceptable, acceptable only in informal contexts, or acceptable in formal contexts

As Figure 5.9 demonstrates, the acceptability of using *different than* with a phrasal complement is split into nearly equal parts among both groups. Roughly one-third of the respondents from each group consider *different than* with a phrasal complement unacceptable while two-thirds find it acceptable at least in some contexts. A comment from one blogger may be reflective of this attitude that recognizes some problems with the traditional rule while still maintaining a preference for upholding it: “‘Different than’ isn’t ‘wrong’ in any absolute sense, but I prefer ‘different from’ in my own writing.”

5.4 SPLIT INFINITIVE

A so-called split infinitive occurs when a modifying adverb is placed between the particle *to* and the base form of a verb, as in (31):

- (31) ...the initiative, which is expected to easily qualify for the November 1996
ballot (news_1995)

According to the traditional rule, splitting an infinitive in this way is considered incorrect. Revising Sentence (31) above to move the adverb after the base form of the verb would be considered preferable according to the traditional rule.

- (32) ...the initiative, which is expected to qualify easily for the November 1996
ballot (news_1995)

The proscription against the split infinitive has been called “a virtual icon of prescriptivism” (Tieken-Boon van Ostade, 2013, p. 4). The resistance to splitting infinitives began in the early to mid-nineteenth century (Tieken-Boon van Ostade, 2013) as a response to the idea that infinitives could not be split in Latin and should therefore not be split in English.

5.4.1 Review of Advice from Usage Guides

There are two general approaches to arguing against the SPLIT INFINITIVE. The first is to point out that Latin infinitive verbs were composed of one word while English infinitive verbs are composed of two (*to* + the base form of the verb). Because English infinitives are made up of two words, there is no reason that a modifying adverb could not be used between the two words.

The other approach is more linguistically informed. It recognizes that the infinitive form of a verb in English is actually a single word (the base form of the verb) and that it is

often accompanied by the particle *to*. As the argument goes, because the infinitive in English is actually one word, inserting a modifying adverb between the *to* and the infinitive is not actually splitting anything. So how could it be wrong?

Regardless of the approach, most of the authors of the usage guides included in this study recognize the illegitimacy of this prescription. Fogarty (2008), whose advice is often on the extreme prescriptive end of the scale takes the opposite approach on this issue, saying “I consider it my calling to dispel the myth that it’s against the rules to split infinitives. It’s fine to split infinitives, and sometimes, I split them when I don’t have to just to maliciously make a point” (p. 55). Strunk and White (2009) and Butterfield (2015) were the most prescriptive in their discussions of the rule; however, neither took a completely prescriptive approach. Butterfield says that

the ban on the split infinitive...has sufficient weight of opinion against it to recommend avoidance when possible, and especially when it is stylistically awkward. But it is neither a major error nor a grammatical blunder, and it is acceptable and at times necessary when considerations of rhythm and clarity call for it. (p. 773)

In total, the split infinitive had the second-lowest prescriptivism index of all of the usage problems, with a rating of 1.95. This indicates that it is widely seen among usage experts as a rule that is not worth following. Yet, because it has become such an icon as Tieken-Boon van Ostade described it, it continues to take up space in US handbooks on usage. None of the usage guides included in this study recommend that this is a rule that should always be followed.

5.4.2 Research Questions and Method of Extracting Data from the Corpora

To study this usage problem in personal blog writing and news writing, I investigated the following research questions:

RQ1_{split Inf.} What proportion of instances of infinitives accompanied by modifying adverbs are split vs. unsplit (i.e., the adverb is placed immediately before or after the infinitive verb phrase) in formal and informal registers of writing?

RQ2_{split Inf.} To what extent do the proportions between these two registers differ?

In order to extract examples of infinitive verb phrases modified by adverbs, I compiled the three search patterns listed in Table 5.7. Simply using the general adverb tag in the CLAWS7 tagset (_RR) would return many unsplit infinitives in which the split alternation would not be possible, as in *to see how* and *to learn more* (cf. **to how see* and **to more learn*). Therefore, the patterns I developed identify as adverbs any word endings in *-ly* in addition to the adverbs *better*, *just*, *still*, *also*, *even*, *further*, *ever*, *always*, *never*, and *not*. I selected these 10 adverbs because they commonly appear in split infinitives in all registers of COCA (Davies 2008).

I then coded the randomly selected instances as outlined in Section 4.2.3.3 to determine whether each instance could reasonably be considered to constitute a split or unsplit infinitive. Because of the way I built my search patterns, false positives like those in (33) and (34) were relatively frequent, accounting for more than a quarter (26%) of the total instances I analyzed.

Table 5.7 Search terms used to extract data from both corpora

Patterns	Regular Expressions	Examples
Split infinitive	<code>r"\bto_\w+[@%]? (?:\w+ly better just still also even furt her ever always never not)_\w+[@%]? \w+_VVI[@%]?"</code>	*It is impossible <u>to entirely separate</u> the gun-ownership from concerns of self-preservation (rightreason_typepad_00004)
Unsplit with adverb after	<code>r"\bto_\w+[@%]? \w+_VVI[@%]? (?:\w+ly better just still also even furt her ever always never not)_\w+[@%]?"</code>	...so I took this as an opportunity <u>to</u> <u>procrastinate further</u> " (matrixarmory_blogspot_00004)
Unsplit with adverb before	<code>r"(?:\w+ly \bbetter \bjust \bstill \balso \beven \bfurther \bever \balways \bnever \bnot)_\w+[@%]? to_\w+[@%]? \w+_VVI[@%]?"</code>	Iran remains determined to get nuclear fuel facilities <u>allegedly to</u> <u>preserve</u> nuclear energy independence" (news_2006)

(33) It seems everyone wants me to like Billy Collins' poetry

(banjo52_blogspot_00023)

(34) ...the logic of computing can assist our human thought processes as we try to

manage daily problems. (thebookshopper_typepad_00005)

After identifying 250 instances of infinitives from both registers, I calculated the proportion of split versus unsplit infinitives in each register.

5.4.3 Results of Corpus Analysis

The results of this analysis are shown in Figure 5.10. These results demonstrate that split infinitives occur often in both registers of writing, though they occur at a higher proportion in personal blog writing.

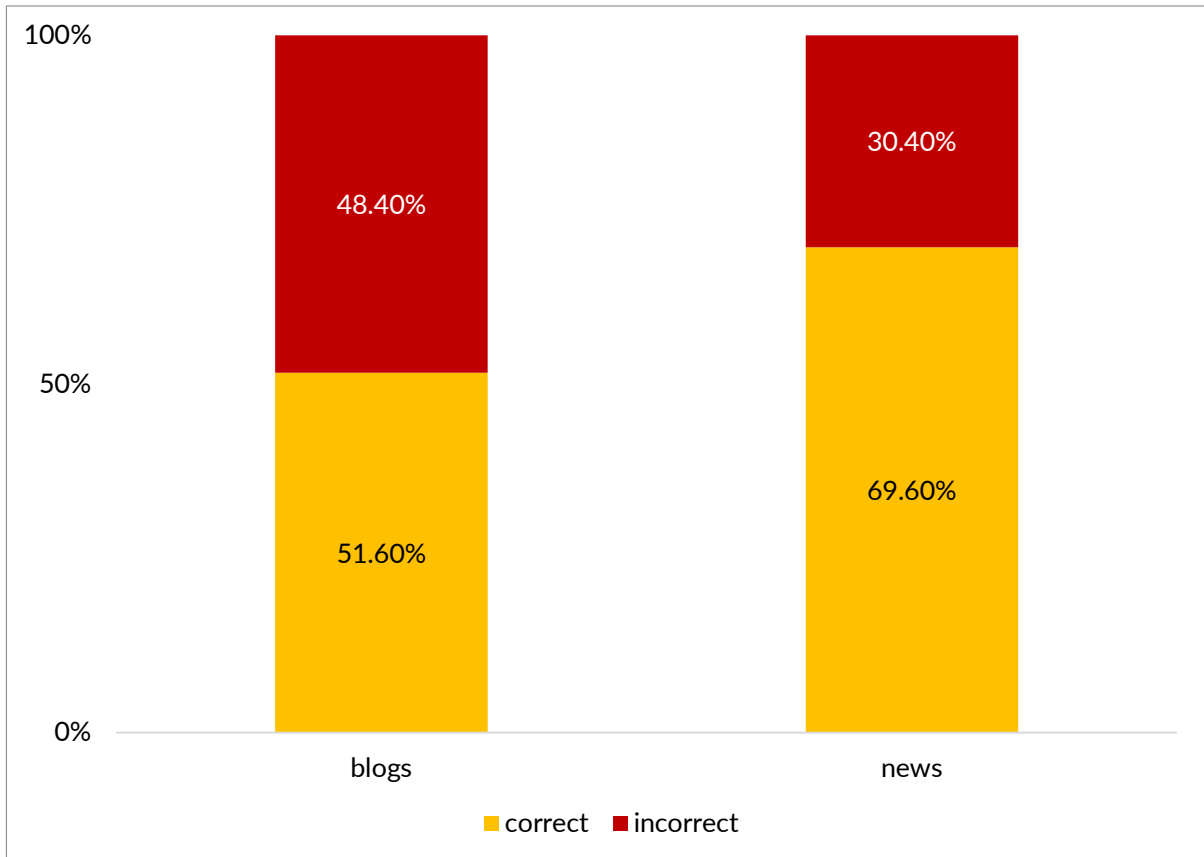


Figure 5.10 Proportion of split and unsplit infinitives in blog and news writing. Split infinitives are marked as incorrect; unsplit infinitives are marked as correct.

5.4.4 Summary of Survey Results

To understand the attitudes related to this usage rule, I asked survey participants whether the following sentence was acceptable or unacceptable, and if acceptable, in which contexts:

(35) *She refused to even think about it.

This sentence flouts the traditional by splitting the verb phrase *to think* with the adverb *even*. According to the prescriptive rule, this sentence would be correctly revised to say “She refused even to think about it.”

Overall, as has been observed in the previous analyses, journalists viewed this usage problem more leniently than bloggers did, though both groups overwhelmingly accepted the

sample sentence as acceptable in at least informal contexts. In fact, 100% of the journalists found the sentence acceptable. Only 12.86% ($n = 9$) bloggers found the sentence unacceptable; none of the journalists did. One possible reason for this finding could be that journalists are echoing the lenient stance the *AP Stylebook* gives on this usage problem: “In many cases, splitting the infinitive or compound forms of a verb is necessary to convey meaning and make a sentence easy to read. Such constructions are acceptable” (verbs, n.d.). Nearly all of the journalists who responded to the survey suggested they were familiar with the *AP Stylebook*. Proportionally much fewer bloggers did, though many bloggers did indicate being familiar with a range of other style guides. Investigating the stance taken by all of the style guides mentioned in the survey was outside the scope of this study, though future research could study the level of prescriptivism in all of those guides to help explain why bloggers might also take very lenient views on the SPLIT INFINITIVE usage problem.

The rule against the split infinitive is widely rejected among usage experts as noted above, and the results from my survey suggest that the same is true for bloggers and journalists. One blogger stated that they are “not hung up on split infinitives, which is the only supposed error I see here.” Another comment from a journalist acknowledged the effect that pronouncements from the *AP Stylebook* has in issues related to linguistic prescriptivism: “I’m assuming the split infinitive is at issue here, but if AP’s fine with it, I am too.”

Interestingly, the fact that journalists seem to be more permissive of splitting infinitives, as shown in Figure 5.11, is not reflected in the usage patterns among news writers, as shown in Figure 5.10. The rule is followed more often in news writing than in blog writing even though a higher percentage of the bloggers who responded to my survey saw the rule as more important to follow than the journalists did.

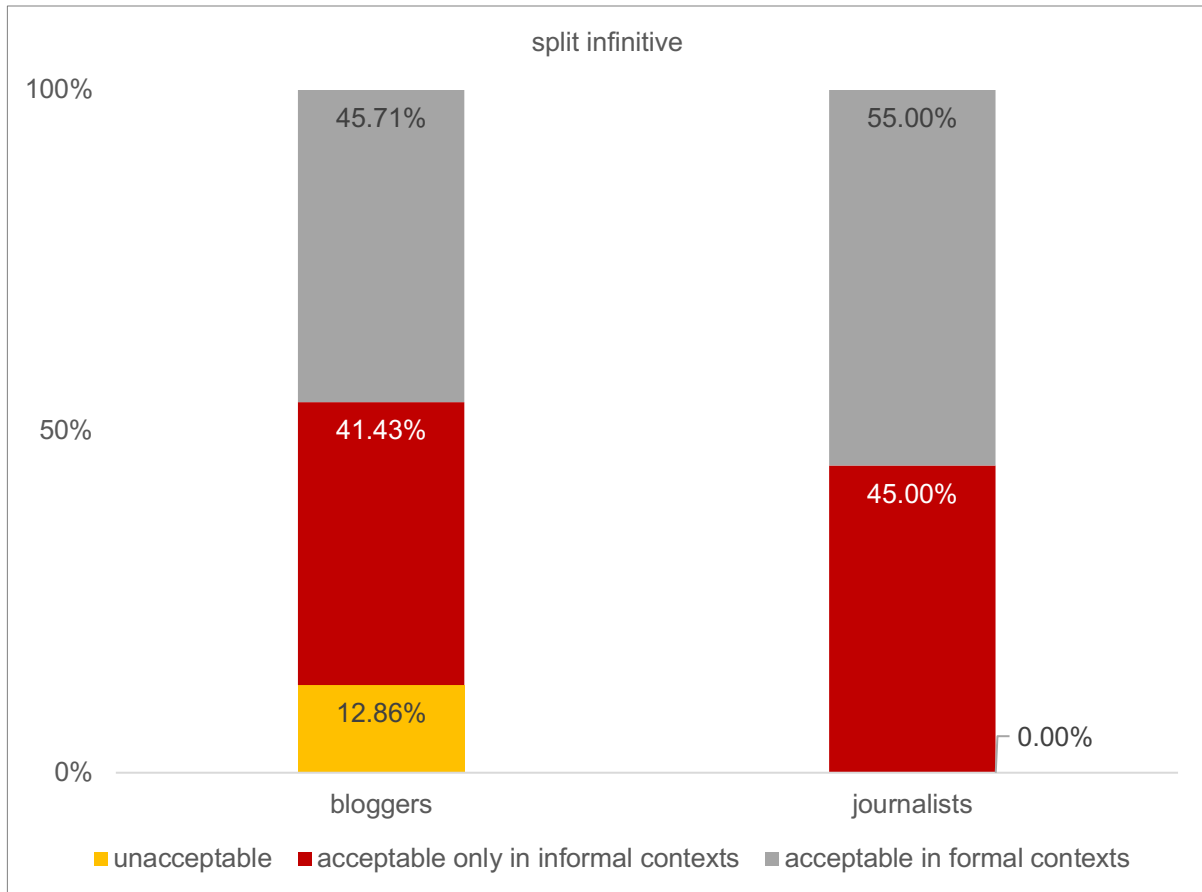


Figure 5.11 Proportion of bloggers and journalists who found the SPLIT INFINITIVE unacceptable, acceptable only in informal contexts, or acceptable in formal contexts

5.5 I FOR ME

The prescription for the I FOR ME usage problem broadly states that nominative pronouns (*I, we, he, she, they*) are correctly used in subject position while accusative pronouns (*me, us, him, her, them*) are correctly used in object position. This rule can be particularly problematic when these pronouns occur in coordinated noun phrases as in the example in (36).

- (36) *Me and Erik were all happy (news_2003)

Because the conjoined noun phrase *Me and Erik* is in the subject position, prescriptivists argue that the pronoun should be in the nominative case, making it become *Erik and I*.

5.5.1 Level of Prescriptivism in Usage Guides

Usage guides address the I FOR ME usage problem under a number of different headings, including “pronouns,” “hypercorrection,” or “between you and I.” In fact, nine of the 11 usage guides included in this study have specific entries with a title of “between you and I” or some derivative. As a result, much of the advice given on how to use coordinate noun phrases is given specifically within the context of this one prepositional phrase.

MWDEU (1994) suggests that using a nominative pronoun where an accusative belongs (such as in the phrase *between you and I*) occurs in “rather more educated varieties of English” while constructions that use an accusative pronoun where a nominative belongs, as in (36) above, are characteristic of “less educated English” (p. 778). As will be shown in the results of this usage problem below, there were no instances of the hypercorrect use of nominative pronouns where accusatives are expected in the data I analyzed for this study. However, there were attested instances of what *MWDEU* calls “less educated” variety.

Of all the usage problems included in this study, the I FOR ME usage problem had the highest overall prescriptivism index at 3.50. More than half of the guides (Batko, 2004; Brians, 2013; Fogarty, 2008; Garner, 2016; O’Connor, 2009; and Strunk & White, 2009) took a highly prescriptive view of this usage problem (see Appendix C), recommending that readers maintain the distinction regardless of context. Both the *American Heritage Guide to Contemporary Usage and Style* (2005) and *MWDEU* (1994) had the lowest scores, with scores that fell exactly in the middle of the scale. According to *MWDEU*, readers

are probably safe in retaining *between you and I* in [their] casual speech, if it exists there naturally, and [they] would be true to life in placing it in the mouths of fictional

characters. But [they] had better avoid it in essays and other works of a discursive nature. It seems to have no place in modern edited prose. (p. 183)

Overall, then, current usage guides highly recommend maintaining this rule.

5.5.2 Research Questions and Method of Extracting Data from the Corpora

To identify the target feature in my corpus analysis, I focused my analysis on what Zwicky (2010) has called “nominative conjoined objects” and “accusative conjoined subjects.” A nominative conjoined object is a coordinate noun phrase in object position in which at least one of the elements is “visibly nominative” (para. 1). An accusative conjoined subject, in contrast, is a coordinate noun phrase in subject position in which at least one of the elements is visibly accusative. These constructions are commonly discussed in entries on the I FOR ME usage problem in usage guides, so they make a useful operationalization of it. Both are prescriptively disparaged. As such, I also included instances of the prescriptively preferred nominative conjoined subjects and accusative conjoined objects in my data to use for comparison.

To determine the extent to which usage of this rule varies in news and blog writing, I investigated the following research questions:

- | | |
|-----------------------|--|
| RQ1 _{IforMe} | What proportion of instances of visibly accusative pronouns appear in coordinate noun phrases in subject and object positions in formal and informal registers of writing? |
| RQ2 _{IforMe} | What proportion of instances of visibly nominative pronouns appear in coordinate noun phrases in subject and object positions in formal and informal registers of writing? |
| RQ2 _{IforMe} | To what extent do the proportions between these two registers differ? |

To find instances of the features, I incorporated the regular expressions shown in Table 5.8 into a Python script and extracted all instances of the text that matched these patterns.

Table 5.8 Search terms used to extract data from both corpora

Patterns	Regular Expressions	Examples
__ and [nominative pronoun]	<code>r"\w+_(?:NP1 NP2 NP) [@%]? (?:and or)_\w+ [@%]? (?:I she he we they)_\w+ [@%]?"</code>	<u>Peter and I</u> were asleep on the sofa (sharon-genealogy_blogspot_00005)
[nominative pronoun] and __	<code>r"(?:\bI \bhe \bshe \bwe \bthey)_\w+ [@%]? (?:and or)_\w+ [@%]? \w+_(?:NP1 NP2 NP) [@%]?"</code>	<u>He and Greenfield</u> began selling to restaurants (news_1992)
[accusative pronoun] and __	<code>r"(?:\bme \bus \bhim \bher \bthem)_\w+ [@%]? (?:and or)_\w+ [@%]? \w+_(?:NP1 NP2 NP) [@%]?"</code>	Both her beautiful brothers and <u>me and Jwan</u> have fallen in love with Lucia (randyreport_blogspot_00015)
__ and [accusative pronoun]	<code>r"\w+_(?:NP1 NP2 NP) [@%]? (?:and or)_\w+ [@%]? (?:me us him them her)_\w+ [@%]?"</code>	...he will win more majors than <u>me and Arnold</u> Palmer combined (news_1999)

I then coded the randomly selected instances as outlined in Section 4.2.3.3 as either a nominative or accusative conjoined noun phrase, and whether it occurred in a subject or an object position.

5.5.3 Results of Corpus Analysis

Nearly three-quarters of the total instances I analyzed (71%) were false positives like those in (37) and (38) below.

- (37) She returns to her father, Lewis Blain Anderson and her mother, Neva Jean Edeburn... (andersenseven_typepad_00002)
- (38) ...have to plot attacks against US and NATO forces in Afghanistan (news_2009)

Because of the high number of false positives in the data, it was not possible to code 1,000 examples as I had hoped to do. Instead, I coded all instances of accusative conjoined noun phrases in the CPBP, of which only 205 were instances of the target feature. This means that the number of accusative conjoined noun phrases that were coded was 45 instances less than the target number of 250. The results of this analysis are shown in Figure 5.12. The overwhelming majority of instances of the I FOR ME usage problem followed the prescriptive rule in both registers. Surprisingly, news writing contained proportionally more instances in which the traditional rule was broken than blog writing. This is likely due to the number of instances that occur in spoken language that is then transcribed and included in the news articles. Because editors are trained to not edit language from interviews, these instances of what prescriptivists may call “bad English” are published in news writing. In addition, the smaller sample of instances from blogs that were coded for this usage problem may have skewed the results to some degree. However, it is unlikely that if 45 additional instances of this usage problem were available to code in the CPBP that the results would have been markedly different.

5.5.4 Summary of Survey Results

To study the attitudes toward this usage problem, survey participants were asked whether the following sentence was acceptable or unacceptable, and if acceptable, in which contexts:

(39) *She told Charles and I the whole story.

This sentence flouts the traditional rule because it uses the nominative pronoun *I* as the object of the verb *told*.

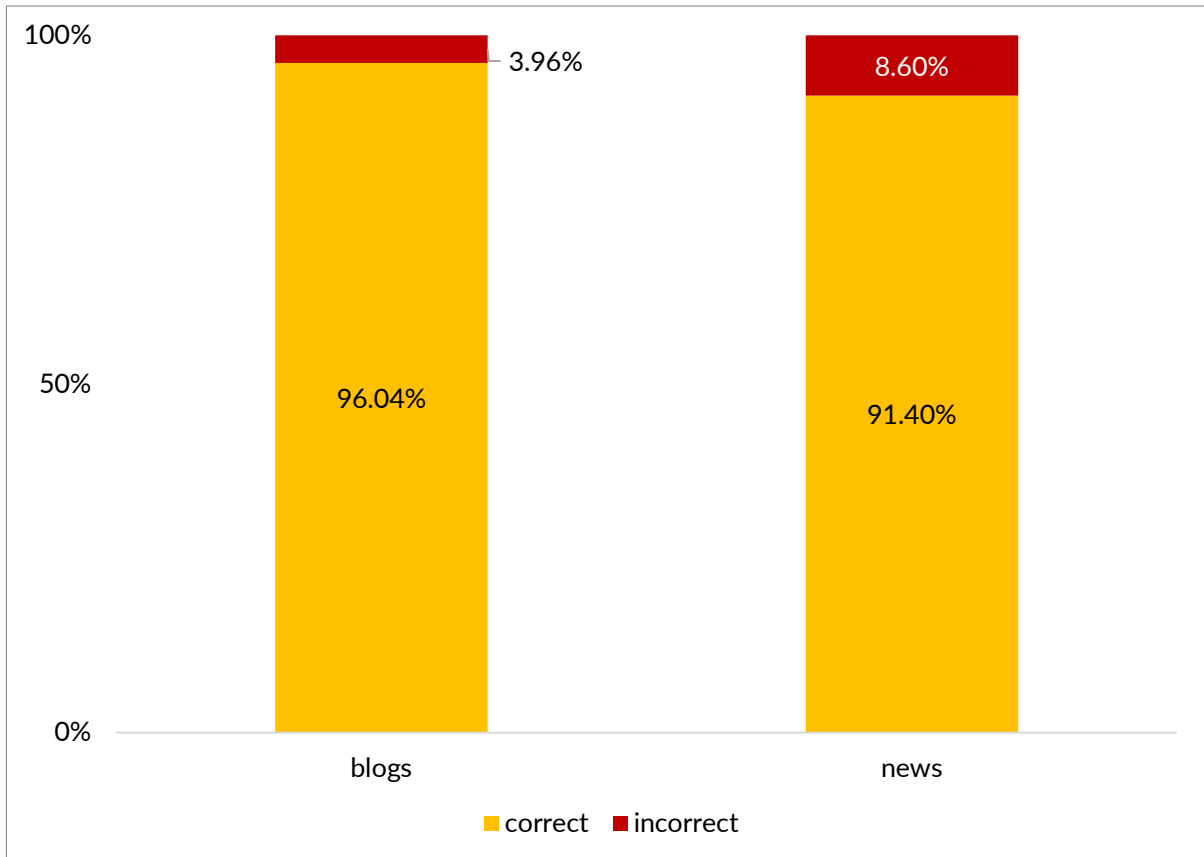


Figure 5.12 Proportion of nominative and accusative conjoined noun phrases that were used correctly and incorrectly in blog and news writing

The attitudes captured in the responses to the question about this usage problem, visualized in Figure 5.13, mirror the guidelines given in the usage guides. That is, a high proportion of both bloggers and journalists find that using nominative pronouns where accusative pronouns are required is unacceptable. A slightly higher proportion of bloggers (73.24%) found Sentence 39 unacceptable, though nearly three-quarters (70.00%) of journalists did as well. Only 8.45% of bloggers and 10.00% of journalists felt that flouting the I FOR ME rule was acceptable in formal contexts.

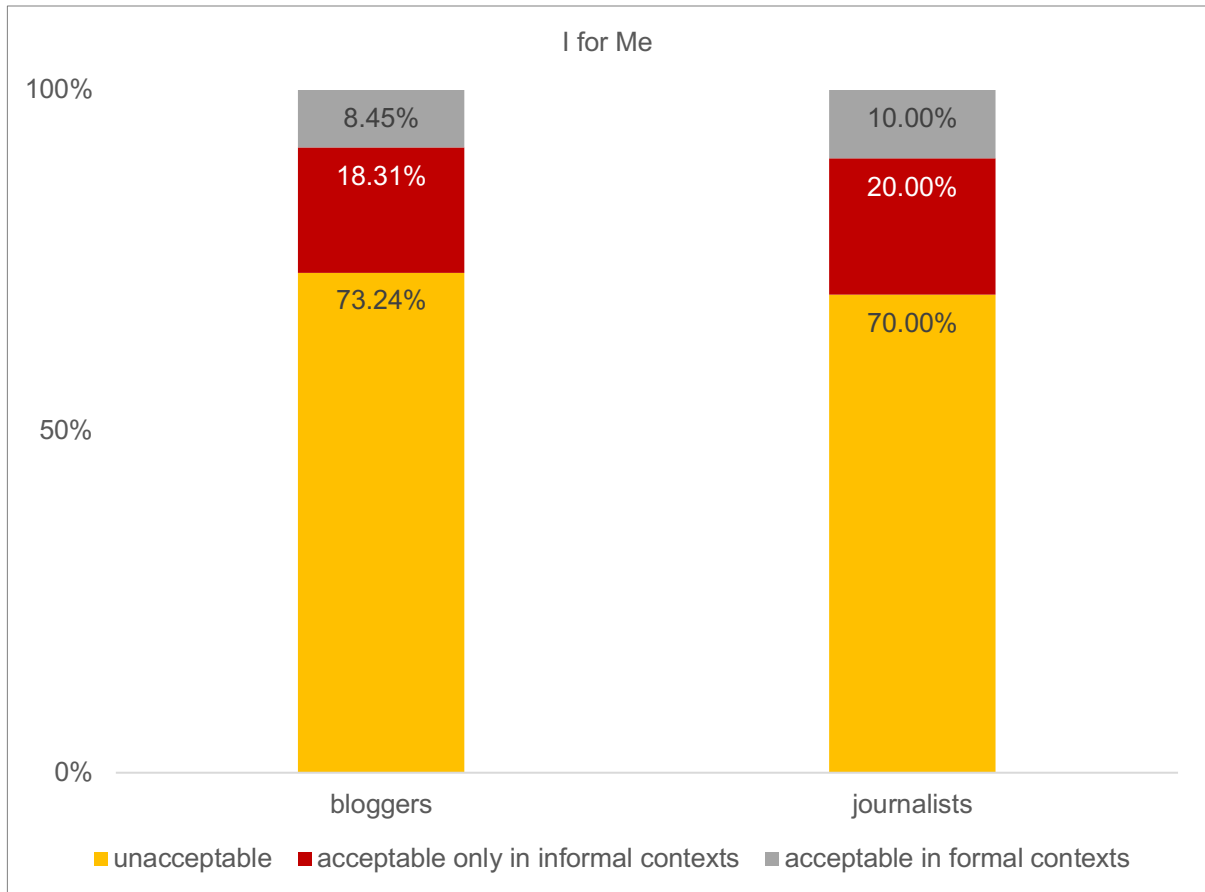


Figure 5.13 Proportion of bloggers and journalists who found the misuse of I FOR ME unacceptable, acceptable only in informal contexts, or acceptable in formal contexts

These results show a high level of linguistic awareness among those in both groups. Using *I* in place of *me* is sometimes considered a hypercorrection—meaning that speakers or writers use *I* in constructions even in coordinate noun phrases in accusative position because to some it sounds more correct or formal even though it is widely shunned as an acceptable alternative to the use of *me*. That bloggers and news writers recognize this use of *I* as problematic and not as overly formal may be evidence of their training on this matter.

5.6 SINGULAR THEY

Though the use of *they* as a singular pronoun has been in use for hundreds of years (the *OED* cites its earliest use as 1375, [“they”, n.d.]), it became a controversial and disparaged variant with the rise of prescriptivism in English in the 18th century (*MWDEU*,

1994). The attitudes of that time are reflected in the traditional rule for singular *they*, which proscribes using *they* to refer to a singular antecedent. There are several options for referring to antecedent nouns whose gender is unmarked and, historically, usage guides have prescribed using *he* to do so, arguing that *he* can function as a gender-neutral singular pronoun. Other common options include *he or she*, *s/he*, *she*, and *they*, with *they* being the most passionately proscribed form.

5.6.1 Level of Prescriptivism in Usage Guides

The prescriptivism index for SINGULAR THEY from the current usage guides included in this study is exactly in the middle of the scale at 2.50, showing that attitudes about the acceptability of SINGULAR THEY are currently in flux. Of all of the usage problems included in this study, the usage guides' treatment of SINGULAR THEY had the widest variation. Only SINGULAR THEY and DIFFERENT TO/THAN/FROM earned ratings that span the entire prescriptivism scale, ranging from a minimum rating of 1 to a maximum rating of 4. The standard deviation for SINGULAR THEY is the largest of any of the usage problems (SD = 1.05, see Appendix C) indicating that opinions about its acceptability vary more than any of the other usage problems.

MWDEU (1994) takes an entirely permissive approach to allowing singular *they*, noting “[t]hey, *their*, *them* are used in both literature and general writing to refer to singular nouns, when those nouns have some notion of plurality about them...Notional agreement is in control, and its dictates must be followed” (p. 903). Peters (2004) takes a similarly tolerant view in allowing singular *they*. She offers evidence from other authorities on usage that “shows that singular use of **they/them/their** after indefinites is now well established in

writing.” She goes on to conclude that “Writers who use singular **they/them/their** are not at fault” (p. 538, boldface in original).

In contrast to *MWDEU* (1994) and Peters (2004), Strunk and White (2009) and O’Connor (2009) take the opposite approach. I quote Strunk and White at length:

Do not use *they* when the antecedent is a distributive expression such as *each, each one, everybody, everyone, many a man*. Use the singular pronoun...A similar fault is the use of the plural pronoun with the antecedent *anybody, somebody, someone*, the intention being either to avoid the awkward *he or she* or to avoid committing oneself to one or the other. Some bashful speakers even say, “a friend of mine told me that they...”

The use of *he* as a pronoun for nouns embracing both genders is a simple, practical convention rooted in the beginnings of the English language. Currently, however, many writers find the use of the generic *he* or *his* to rename indefinite antecedents limiting or offensive. Substituting *he or she* in its place is the logical thing to do if it works. But it often doesn’t work, if only because repetition makes it sounding boring or silly. (pp. 60–61)

O’Connor’s (2009) approach is similarly prescriptive in advising readers that “[s]trictly speaking, one person can’t be a *they*. Yes, it’s tempting to use *they* and *them* when you don’t know whether the somebody is a he or a she. But resist the temptation” (p. 14).

The entries from Garner (2016) and Fogarty (2008) each received a prescriptivism index of 2.5. Fogarty admits that when she is writing a formal document, she avoids singular *they*. But she also acknowledges that she believes singular *they* will eventually become acceptable. However, she cautions,

it takes a bold, confident, and possibly reckless person to use *they* with a singular antecedent today....[I]f you are a respected editor in charge of writing a style guide for your entire organization, you can get away with making it acceptable to use *they* with a singular antecedent. I would even encourage you to do so, and there are a variety of credible references that will back you up including the *Random House Dictionary* and *Fowler's Modern English Usage*. You would be in the company of revered authors such as Jane Austen, Lewis Carroll, and Shakespeare. But, if you are responsible to superiors, there's a good chance that at least one of them will think you are careless or ignorant if you use *they* with a singular antecedent.

So here's the quick and dirty tip: rewrite your sentences to avoid the problem. If that's not possible, check if the people you are writing for have a style guide. If not, use *he or she* if you want to play it safe, or use *they* if you feel bold and prepared to defend yourself. (p. 61)

These excerpts from usage-guide entries on SINGULAR THEY demonstrate the range of opinions that guide authors espouse on the subject.

5.6.2 Research Questions and Method of Extracting Data from the Corpora

To determine the extent to which SINGULAR THEY is used in both news and personal blog writing, I investigated the following research questions:

RQ1_{sgThey} What proportion of instances of singular pronouns or combinations of singular pronouns (e.g., *he or she*) are used to refer to indefinite pronominal antecedents in formal and informal registers of writing?

RQ2_{sgThey} What proportion of instances of plural pronouns are used to refer to indefinite pronominal antecedents in formal and informal registers of writing?

RQ3_{sgThey} To what extent do the proportions between these two registers differ?

While singular and plural pronouns are often used to refer to indefinite gender-neutral nouns (e.g., *partner*, *teacher*, *child*, etc.), I restricted my analysis to include only instances of pronouns that have indefinite pronouns as their antecedents. Doing so allowed me to more narrowly operationalize my data. I had originally written a script that matched instances of common gender-neutral singular nouns identified by Biber (2006); however, analyzing the data in this way resulted in many false positives. Focusing the analysis on only those pronouns that occurred in the vicinity of indefinite pronouns also resulted in many false positives, but it likely allowed for a more focused, precise analysis. In addition, some of the usage guides (e.g., Butterfield, 2015; Garner, 2016, *MWDEU*, 1994; and Peters, 2004) include sections that specifically address the issue of pronominal reference to indefinite pronoun antecedents. For these reasons, I focused my analysis on the use of SINGULAR THEY with reference to the indefinite pronouns *anyone*, *anybody*, *everyone*, *everybody*, *no one*, *nobody*, *someone*, and *somebody*.

Of all the usage problems included in this study, SINGULAR THEY can occur in the widest variety of grammatical constructions, so it poses a challenge methodologically. In order to capture instances of SINGULAR THEY and compare them to instances where other pronouns were used to refer to indefinite pronouns, I developed the framework in Table 5.9 that divides the different pronouns into categories based on number and case. I also

considered possible combinations that people use to refer to gender-neutral singular antecedents such as *his or her* or *s/he*.

Table 5.9 Third-person pronouns to consider when studying variation in singular *they*.

	Subject	Object	Possessive	Reflexive
Masculine	he (not followed by “or she” and not preceded by “she or”)	him (not followed by “or her” or “or herself” and not preceded by “her or”)	his (not followed by “or her” or “or hers” or “or herself” and not preceded by “her or” or “hers or”)	himself (not followed by “or himself” and not preceded by “herself or”)
Feminine	she (not followed by “or he” and not preceded by “he or”)	her (not followed by “or him” or “or himself” or preceded by “him or” or “his or”)	her (not followed by “or him” or “or himself” and not preceded by “him or” or “his or”) hers (not followed by “or his” and not preceded by “his or”)	herself (not followed by “or herself” and not preceded by “himself or”)
Indefinite	one	one	one’s	oneself
Combined	s/he he or she she or he he/she she/he	him or her her or him him/her her/him	his or her his or hers her or his hers or his his/her his/hers her/his hers/his	his or herself her or himself him or herself himself or herself herself or himself himself/herself herself/himself
Plural	they	them	their theirs	themselves themselves

To find instances of SINGULAR THEY in both corpora, I incorporated regular expressions written in Python to find instances of the pronouns in Table 5.9 only when preceded by an indefinite pronoun in the same sentence or in the preceding sentence. (The regular expressions I used to find each of these patterns are shown in Appendix E.) Previous research (Balhorn, 2009) used an eight-word threshold; however, lengthening this threshold to the instances contained in the same sentence or the preceding sentence allowed me to capture more instances of pronominal references to indefinite pronouns. Because the data in the news corpus was not structured exactly the same way as the data in the CPBP, I extracted

only instances of a pronoun with an indefinite pronoun in the immediate line or in the previous line in the text file. As noted in Section 4.2.1, these lines did not always constitute complete sentences.

In order to extract the lines with only masculine and feminine pronouns that appeared alone and not in combined constructions (see Table 5.9), I wrote the script to temporarily replace instances of combined forms (e.g., *he or she*, *her or him*, etc.) with filler text so that the script would match only instances that were examples of masculine or feminine pronouns alone. I then ran a script that extracted all of the combined, indefinite, and plural pronouns listed in Table 5.9 that contained an indefinite pronoun in the preceding text.

I coded each KWIC line to determine if the pronoun referred to the indefinite pronoun in the pretext or to some other noun. In spite of the specificity and care I took in developing the programs to extract relevant data, there was still a high number of false positives in the data returned. More than two-thirds (69%) were false positives. Sentence (40) shows an example of a false positive returned in the data. Sentence (41) in contrast shows an example in which *they* is used as a proform to refer to the indefinite pronoun *no one*. In both examples, the term identified for analysis appears underlined and in boldface. The antecedent for both appears underlined and a potential intervening false antecedent appears in italics.

- (40) Sadly, for **many of our young people** whose lives have been impacted by foster care, when they emancipate from foster care, *no one* is even making sure **they** have a hot meal (outofthefostercarebox_wordpress_00034)
- (41) I have absolutely no desire to be white, and **no one** I know sits down in a *stylist's* chair because **they** want to be white (news_2009)

I coded each line in the data to determine whether the pronoun referred to an indefinite pronoun or to some other noun in the sentence. For instances in which the antecedent was an indefinite pronoun, I recorded the indefinite pronoun and noted whether the pronoun used was singular or plural. I considered combined forms such as *he or she* as well as the indefinite pronoun *one* and its derivatives as singular. I then marked the lines that contained singular proforms as correct and those that contained plural proforms as incorrect.

5.6.3 Results of Corpus Analysis

The results of this analysis are shown in Figure 5.14. The most striking observation about the data visualized in this figure is the proportion of times that the prescriptive rule for singular *they* is violated in both registers. Roughly 80% of references to an indefinite pronoun were made with a form of *they* in both registers. Conversely, only approximately 20% of references to an indefinite pronoun were made with either a singular pronoun, a combination of singular pronouns, or the indefinite pronoun *one* in both registers. Of the usage problems included in the current analysis, *singular they* is flouted proportionally more often than any other, including the *split infinitives*, which received a lower prescriptivism index rating.

5.6.4 Summary of Survey Results

Results from the survey data show greater variance in the attitudes that both groups have regarding the use of singular *they*. Survey participants were asked whether the following sentence was acceptable or unacceptable, and if acceptable, in which contexts:

(42) ***Everyone** has their own style.

This sentence flouts the traditional rule because it uses the plural *they* to refer to the singular antecedent *everyone*.

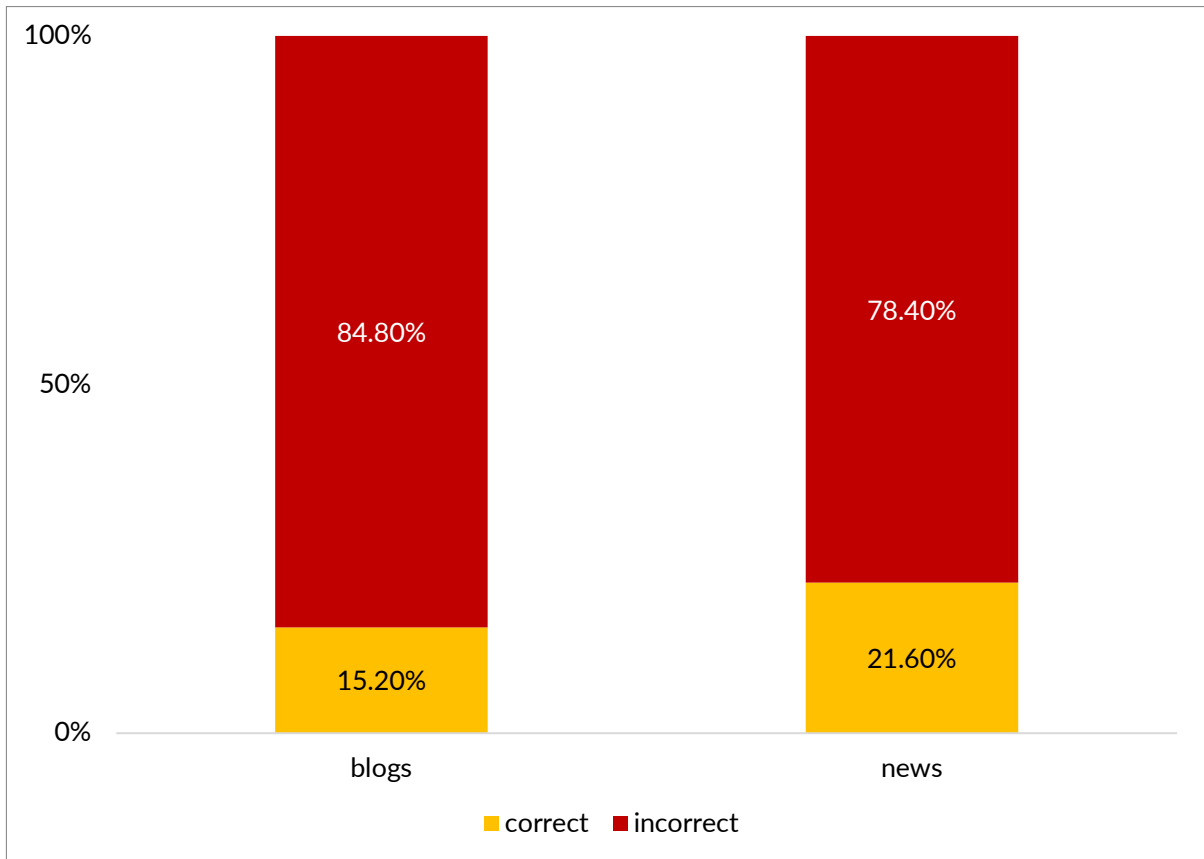


Figure 5.14 Proportion of singular and plural pronouns that reference an indefinite pronoun antecedent in personal blog writing and news writing. Instances in which a singular pronoun was used to refer to an indefinite pronoun were marked as correct; instances in which a plural pronoun was used to refer to an indefinite pronoun were marked as incorrect.

Journalists showed a greater acceptance overall for this use of SINGULAR THEY. Of the journalists who participated in the survey, 85% found this use of SINGULAR THEY acceptable in some contexts with just under two-thirds (60.00%) accepting the use even in formal contexts. Proportionally fewer bloggers (51.43%) accepted SINGULAR THEY in formal contexts, though the acceptability profiles of both, as shown in Figure 5.15 are quite similar.

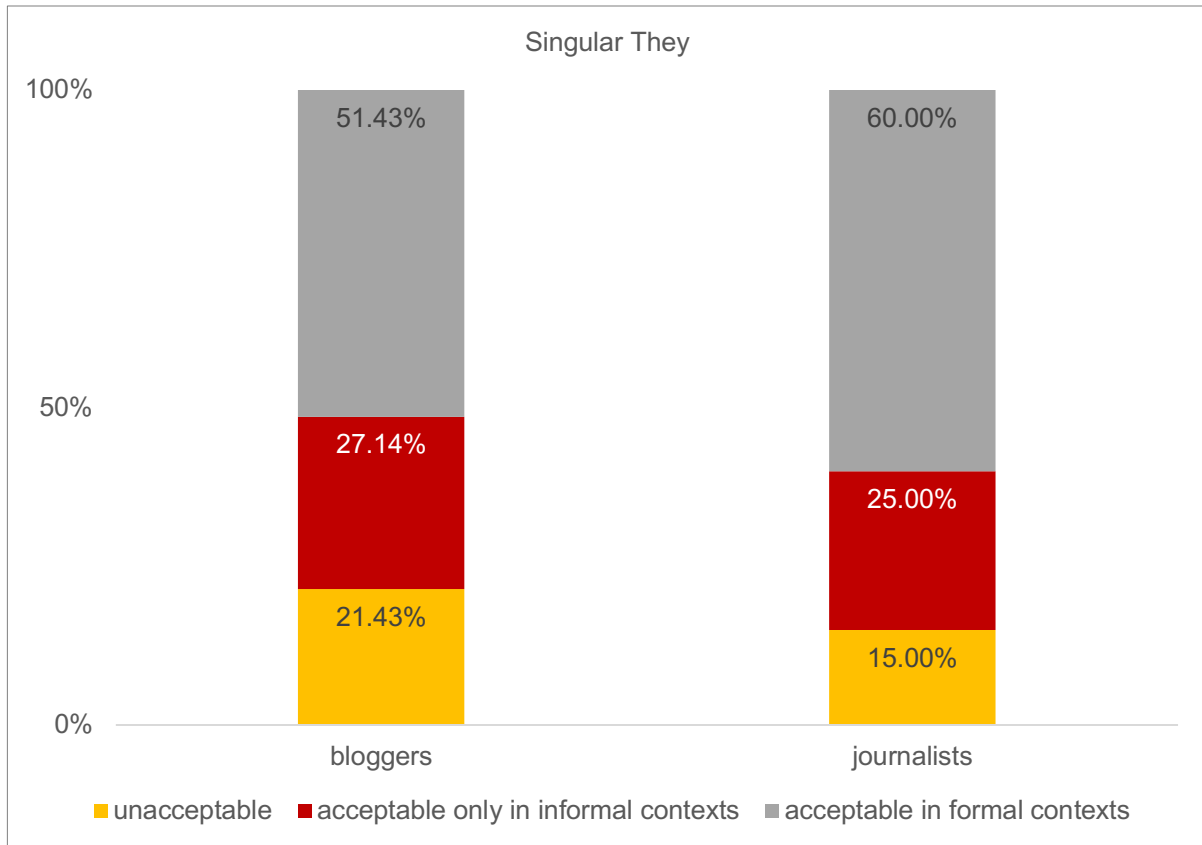


Figure 5.15 Proportion of bloggers and journalists who found *they* used as a singular pronoun unacceptable, acceptable only in informal contexts, or acceptable in formal contexts

Comparing the charts in Figure 5.14 and Figure 5.15, we can observe that the proportion of instances in which the rule is followed roughly corresponds with the proportion of people who find breaking the rule unacceptable. In other words, roughly 20% of respondents to the survey find flouting the rule unacceptable and roughly 20% of all instances coded in the data followed the rule. This apparent correlation indicates that the views of bloggers and journalists are reflected to a similar extent in writing from both groups. Comparing this data with the middling prescriptivism index of 2.5 for this usage problem indicates that usage guides are generally more prescriptive than bloggers and journalists when it comes to SINGULAR THEY.

The following comment left by a blogger who participated in the survey succinctly encapsulates the changing attitudes toward *singular they* and illuminates some possible reasons why the form is so common in both registers of writing even though some still find it unacceptable.

Ugh. The gender neutral possessive is in a state of flux, and many formal arbiters of style now say that this formerly taboo usage is allowed and preferred. The conflict between singular and plural sets my teeth on edge, but it is certainly more concise than “his or her.” I don't really like this one, but I am trying to change with the times and accept it.

5.7 LESS/FEWER

The prescription for the LESS/FEWER usage problem states that *fewer* is correctly used with plural countable nouns (e.g., *apples*, *chairs*, *items*) and *less* is correctly used with uncountable nouns (e.g. *milk*, *attention*, *patience*).

5.7.1 Level of Prescriptivism in Usage Guides

Current usage guides generally recommend upholding this distinction as evidenced by its prescriptivism index of 3.32. No guides included in this study took an entirely lenient approach—that is, none was found to have a prescriptivism index of 1. The minimum rating for any guide in the collection was 2.5; the highest was 4 (see Appendix C). Only the NONE IN PLURAL CONTEXT usage problem had a range narrower than LESS/FEWER.

Peters (2004), whose entry on LESS/FEWER was the most lenient of any of the guides, calls the decision between *fewer* and *less* “a stylistic choice, between the more formal **fewer** and the more spontaneous **less**. **Fewer** draws attention to itself, whereas **less** shifts the focus on to its more significant neighbors” (p. 205, boldface in original). *MWDEU* (1994) says that

readers who are native speakers of English can trust their ear when making a choice between *less* and *fewer*. Brians (2013), on the other hand, advises readers to actively “learn the difference” between these two words so they can “avoid [the] ire” (p. 15) of people who still make the distinction. Garner’s (2016) advice is more moderate, advising readers to preserve the distinction while also acknowledging that “[t]he linguistic hegemony by which *less* has encroached on *fewer*’s territory is probably now irreversible” (p. 560).

5.7.2 Research Questions and Method of Extracting Data from the Corpora

In order to study the ways that *less* and *fewer* are used in actual usage, I investigated the following research questions:

- RQ1_{*less/fewer*} What proportion of instances of *fewer* are used with plural countable nouns versus uncountable nouns in formal and informal registers of writing?
- RQ2_{*less/fewer*} What proportion of instances of *less* are used with plural countable nouns versus uncountable nouns in formal and informal registers of writing?
- RQ3_{*less/fewer*} To what extent do the proportions between these two registers differ?

To extract instances of these features, I incorporated the regular expressions shown in Table 5.10 into a Python script to find instances of *less* and *fewer* immediately followed by a noun. The CPBP contained on 175 usable instances of patterns with *less*.

Table 5.10 Search terms used to extract data from both corpora

Patterns	Regular Expressions	Examples
less	<code>r"\bless_\w+[@%]? \w+_N\w+[@%]?"</code>	...and yet, there's <u>less summer</u> than ever before... (news_2010)
fewer	<code>r"\bfewer_\w+[@%]? \w+_N\w+[@%]?"</code>	Be careful, plan ahead and make <u>fewer mistakes</u> . (diysilkscreenprinting_blogspot_00008)

After collecting the data, I coded each noun immediately following *less* and *fewer* as either countable or uncountable. I then marked each line as correct or incorrect according to the following logic:

- Lines in which a countable noun followed *fewer* were marked correct.
- Lines in which an uncountable noun followed *fewer* were marked incorrect.
- Lines in which a countable noun followed *less* were marked incorrect. However, lines that were included after the phrase *one less* were marked as correct as supported by Garner (2016), *American Heritage* (2005), and O'Connor (2009).
- Lines in which an uncountable noun followed *less* were considered correct.

5.7.3 Results of Corpus Analysis

The results for this analysis are shown in Figure 5.16. The data visualized in the figure show that this rule is observed to a great extent in both registers, though the distinction appears to be slightly more carefully maintained in news writing than in blog writing.

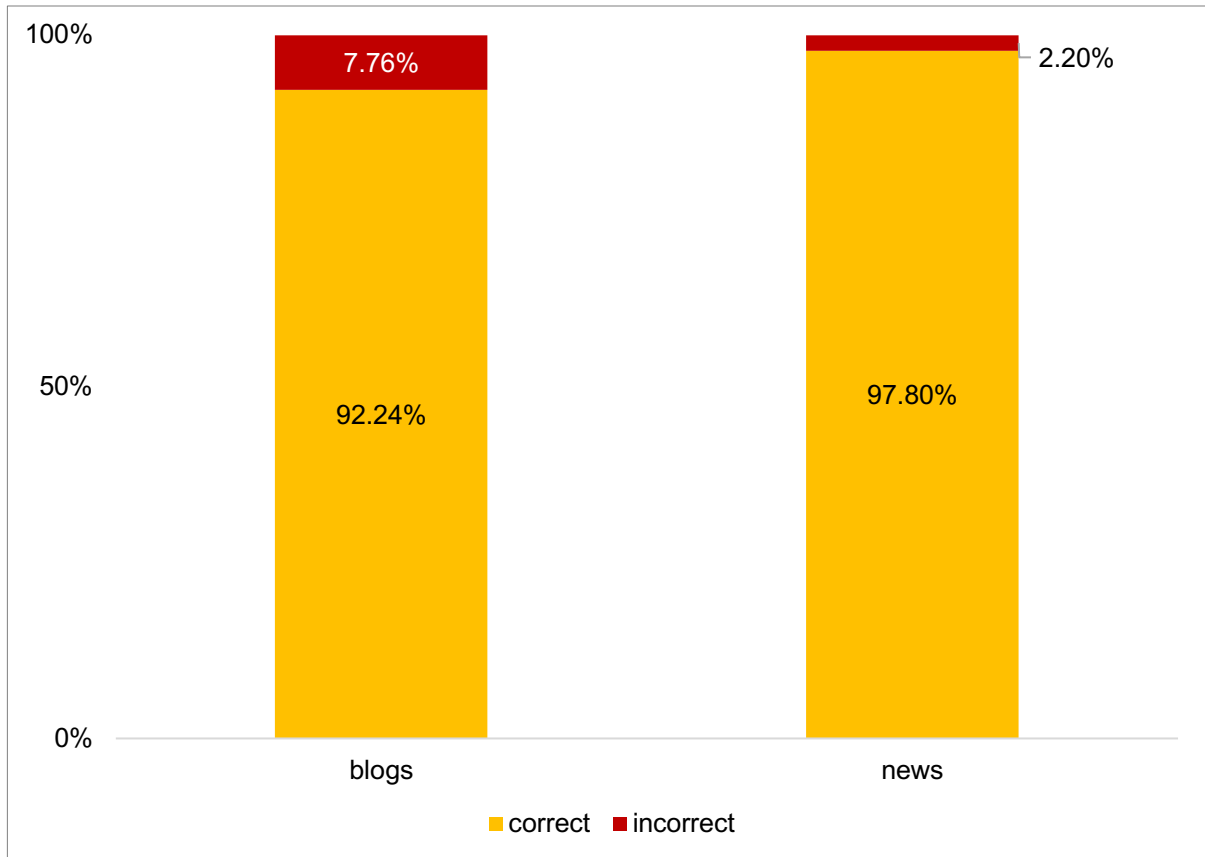


Figure 5.16 Proportion of *less* and *fewer* used to modify plural countable nouns in blog writing and news writing. Instances in which *less* was used to modify a plural countable noun were marked as incorrect; instances in which *fewer* was used to modify a plural countable noun were marked as correct.

5.7.4 Summary of Survey Results

Survey participants were asked whether the following sentence was acceptable or unacceptable, and if acceptable, in which contexts:

- (43) *There were less accidents this year than last year.

This sentence flouts the traditional rule by using *less* to modify the plural countable noun *accidents*.

The majority of respondents in both groups considered the sentence that flouts this rule to be unacceptable in any context. Roughly one-third of the respondents accepted this usage in certain contexts.

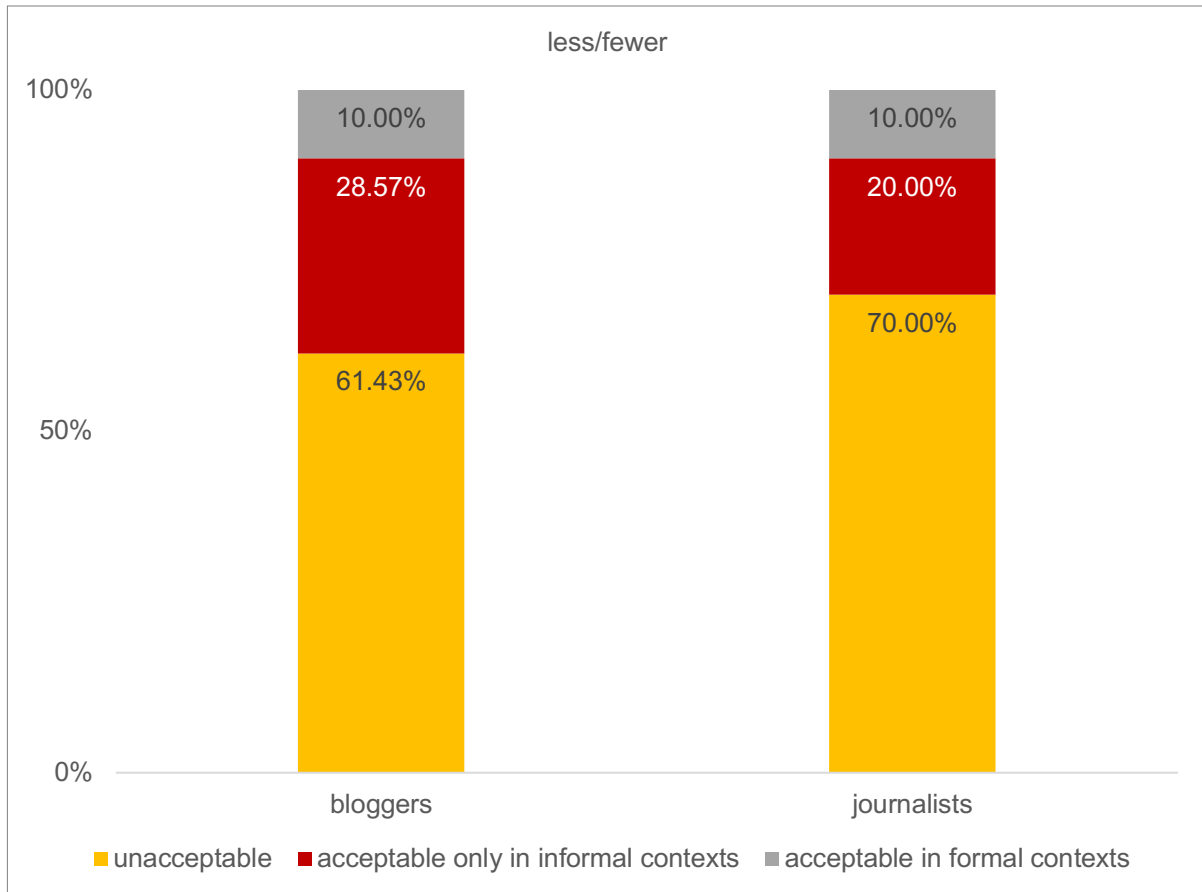


Figure 5.17 Proportion of bloggers and journalists who found the misuse of LESS/FEWER unacceptable, acceptable only in informal contexts, or acceptable in formal contexts

Comparing the proportion of participants who found this usage unacceptable (the yellow portions of the bars in Figure 5.17) to the proportion of instances in which the distinction is maintained in actual usage (the yellow portions of the bars in Figure 5.16), we can observe that the former is substantially less than the latter. This observation suggests that there are potentially many writers who consciously take a lenient position on the rule while still maintaining the prescriptive distinction in their own writing. A similar observation can be seen with LAY/LIE, WHO/WHOM, SPLIT INFINITIVE, and NONE IN PLURAL CONTEXT. Because the surveys were anonymously gathered, it is not possible to confirm this hypothesis by mapping different users' survey responses to the actual text they produced in the corpora.

However, future studies might take up this investigation in order to establish such a connection more firmly.

Responses from survey participants ranged from those who took a highly prescriptive stance with this usage problem (one participant seemed to be shouting their opinion of “FEWER, DAMMIT, FEWER”), to those who were not certain of which is acceptable (another respondent hedged heavily, saying “seems like ‘fewer’ may be a better choice, not sure.”), to those who willingly accepted *less* in place of *fewer* in certain contexts (“So technically I know it should be fewer, but I know what the speaker is saying and in informal conversation wouldn’t judge”).

5.8 NONE IN PLURAL CONTEXT

The issue with words like *none* and *each* stems from the fact that they are grammatically singular, but they are often used semantically to refer to multiple things. For example, if “no one” or “none” of the cookies are on the table, the presupposition is that there are multiple cookies to begin with, and all of the cookies are not on the table (see Biber et al., 1999, pp. 184–185 for a discussion of concord patterns with *none*).

In its strictest sense, the prescription for the NONE IN PLURAL CONTEXT usage problem states that *none* is singular and should therefore agree with singular verb forms. However, grammarians and usage experts have accepted the plural use of *none* since at least the late nineteenth century. Hall’s (1917) *English Usage: Studies in the History and Uses of English Words and Phrases* quotes the English grammarian John Nesfield who in 1898 said that *none* “was originally used only as a Singular...the plural sense is now equally or more common” (p. 39).

5.8.1 Level of Prescriptivism in Usage Guides

More than 100 years after Nesfield's pronouncement, Batko (2004) wrote that the use of *none* with singular verbs was "still in flux." She goes on to describe the rule in somewhat complicated terms:

Unlike the indefinite pronouns listed earlier, "none," along with "some," "any," and "all," isn't always singular. It can be either singular or plural depending on the meaning of your sentence. The rule of thumb is: If it refers to a group in total, or to a thing as a whole, then "none" is usually viewed as singular and takes a singular verb. If it refers to a number of things, meaning your emphasis is on the quantity of the parts and not on the whole, then "none" is considered plural. (p. 114)

Garner (2016) says that the singular or plural forms are both correct. The singular form (e.g., *none is*) adds emphasis to the idea being expressed, but he notes that it can also sound "stilted" (p. 629).

Of all the usage guides included in this study, the NONE IN PLURAL CONTEXT usage problem received the lowest overall prescriptivism rating with an index of 1.45. It also had the smallest standard deviation ($SD = 0.5$) and the lowest maximum score ($max = 2$). Considered together, these descriptive statistics illustrate that current usage guides no longer require *none* to be used exclusively as a singular pronoun.

5.8.2 Research Questions and Method of Extracting Data from the Corpora

In order to better understand how *none* is used in formal and informal writing, I compared the different ways that singular and plural uses of *none* appeared in both registers. Specifically, I addressed the following research questions:

- RQ1_{none} What proportion of instances of *none* can be observed to agree with singular and plural verb forms in formal and informal registers of writing?
- RQ2_{none} What proportion of instances of *none* that appear with singular verb forms also occur with singular nouns in formal and informal writing?
- RQ3_{none} What proportion of instance of *none* that appear with plural verb forms also occur with plural nouns in formal and informal writing?
- RQ4_{none} To what extent do these proportions between these two registers differ?

To identify the target feature, I included the regular expressions shown in Table 5.11 into a Python script.

Table 5.11 Search terms used to extract data from both corproa

Patterns	Regular Expressions	Example
none is/are	<code>r"\bnone_\w+[@%]? \w+_(:VBDR VBDZ VBZ VBR VD0 VDZ VH0 VHZ VV0 VVZ)[@%]?"</code>	...but <u>none were</u> harmed, thank goodness" (twiceremembered_blogspot_00005)
none of the ___ is/are	<code>r"\bnone_\w+[@%]? of_\w+[@%]? \w+_NN\w* \w+_(:VBDR VBDZ VBZ VBR VD0 VDZ VH0 VHZ VV0 VVZ)[@%]?"</code>	<u>None of the buildings is</u> an architectural masterpiece" (news_2001)
none of ___ is/are	<code>r"\bnone_\w+[@%]? of_\w+[@%]? \w+_ \w+[@%]? \w+_(:VBDR VBDZ VBZ VBR VD0 VDZ VH0 VHZ VV0 VVZ)[@%]?"</code>	...with three board members, <u>none of whom seem</u> to have any real expertise in education (brainsandeggs_blogspot_00011)
there is/are none	<code>r"\bthere_\w+[@%]? \w+_VB\w+[@%]? none_\w+[@%]?"</code>	Now <u>there are none</u> (news_2004)

The first pattern matched instances of *none* followed by *were*, *was*, *is*, *are*, *do*, *does*, *have*, *has*, the base form of a lexical verb, or the -s form of a lexical verb. I selected these verbs because they are marked for number, so they allowed me to determine whether the verb

form is plural or singular. The second regular expression matched instances that follow the pattern “none of the [noun]” followed by one of the verbs listed above. The third pattern matched instances of “none of [noun]” followed by instances of the verbs listed above. And the fourth pattern matched instances of *none* in existential *there* constructions formed with a *be* verb. Table 5.11 includes example sentences matched by each of the four patterns.

After the patterns were extracted, I coded the verb form as either singular or plural. The antecedent noun for each instance was also identified and its number was coded as well. This step was important because it allowed me to determine if the antecedent noun of *none* was singular or plural. Comparing the number of the antecedent noun to the number of the verb occurring with *none* allowed me to determine if the number of the antecedent noun and the number of the verb form matched. In cases where the number of the antecedent noun phrase and the number of the verb phrase matched, I marked the line as “match.” In cases where they did not match, I marked the line as “no match.” For example, consider Sentence (44).

- (44) ...and whole lawns and farm fields have been dug up in search of *bones and bone fragments* though none were found (etb-history-theology_blogspot_00008)

This sentence uses the plural verb form *were* with the pronoun *none*. The antecedent of *none* in this sentence was identified as *bones and bone fragments*. The plural nature of this antecedent noun phrase corresponds to the plural verb form *were*, and so this line would be marked as a match.

5.8.3 Results of Corpus Analysis

Before presenting the results of this analysis, it is critical to discuss an important limitation with the method I adopted to study this usage problem. In some cases, a writer might intentionally use a verb form that does not match the number of the noun phrase in order to alter the emphasis of the sentence. For example, if Sentence (44) above were rephrased with a singular verb form as in Sentence (45) below, the grammar is not incorrect even by prescriptive standards—in fact, a highly conservative prescriptivist may even argue that the use of the singular verb in (45) is more correct than the use of the plural verb form in (44). Using a singular verb form has the rhetorical effect of placing emphasis more directly on the fact that not a single bone or fragment was found.

- (45) ...and whole lawns and farm fields have been dug up in search of *bones and bone fragments* though none was found (etb-history-theology_blogspot_00008)

Conducting an analysis that reliably accounts for the intended meaning of each instance of *none* is outside the scope of this study, as doing so would require asking the authors of the texts directly what their intended meaning was. As a result of these limitations, the analysis presented here avoids the use of the terms “correct” and “incorrect” used in the previous analyses. Instead, I use the words “match” and “no match” to label each instance of the verb form and antecedent nouns accompanying the pronoun *none*. By doing so, I avoid labeling as “incorrect” instances where the author intentionally selected the verb form in order to convey a deliberate meaning while offering some insight into this usage problem by identifying the proportion of numerical agreement between verbs following pronominal *none* and the nouns used as its antecedents.

Figure 5.18 shows the proportion of times the number of the verb form accompanying *none* matched the number of the antecedent noun and the proportion of times it did not across both registers. In both registers, the proportion of times that the number of the verb matched the number of the antecedent noun outweighed the proportion of times that they did not match. Though in both registers, the proportion of non-matches was relatively high as well.

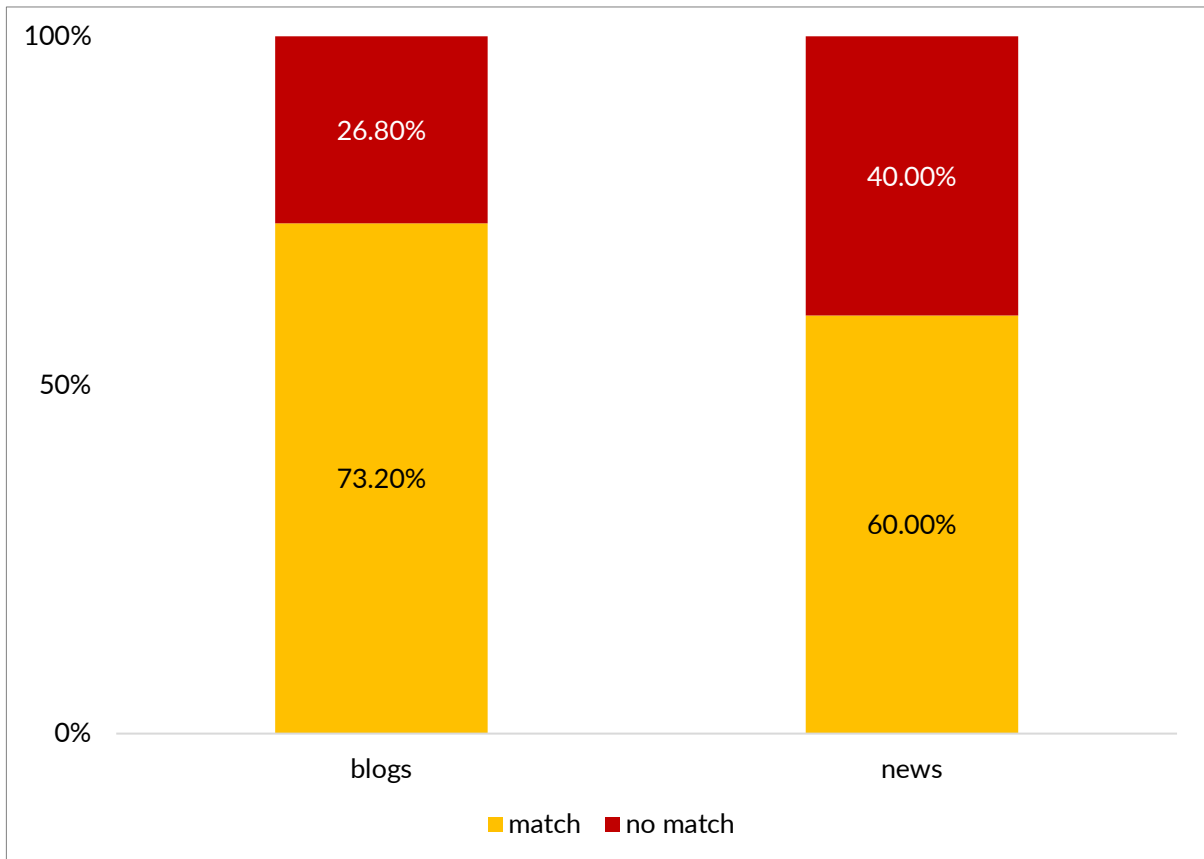


Figure 5.18 Proportion of *none* used in singular or plural contexts in blog and news writing. Instances in which the number of the verb form accompanying *none* matches the number of the antecedent noun are marked as a match. Instances in which the number of the verb and the noun do not match are marked as no match.

5.8.4 Summary of Survey Results

Survey participants were asked whether the following sentence was acceptable or unacceptable, and if acceptable, in which contexts:

- (46) None were left on the table.

The data in Figure 5.19 shows the proportion responses to the question of whether participants found the usage acceptable or unacceptable. Considering the fact that current usage guides are essentially in agreement that *none* can be used as a singular or a plural pronoun, it is surprising to see such a high proportion of respondents who feel that the use of *none* in the sample sentence given in the survey was unacceptable in all instances. One blogger summarized this view, stating that “‘were’ is plural; ‘None’ is singular.” A journalist similarly commented briefly that “None takes a singular verb.”

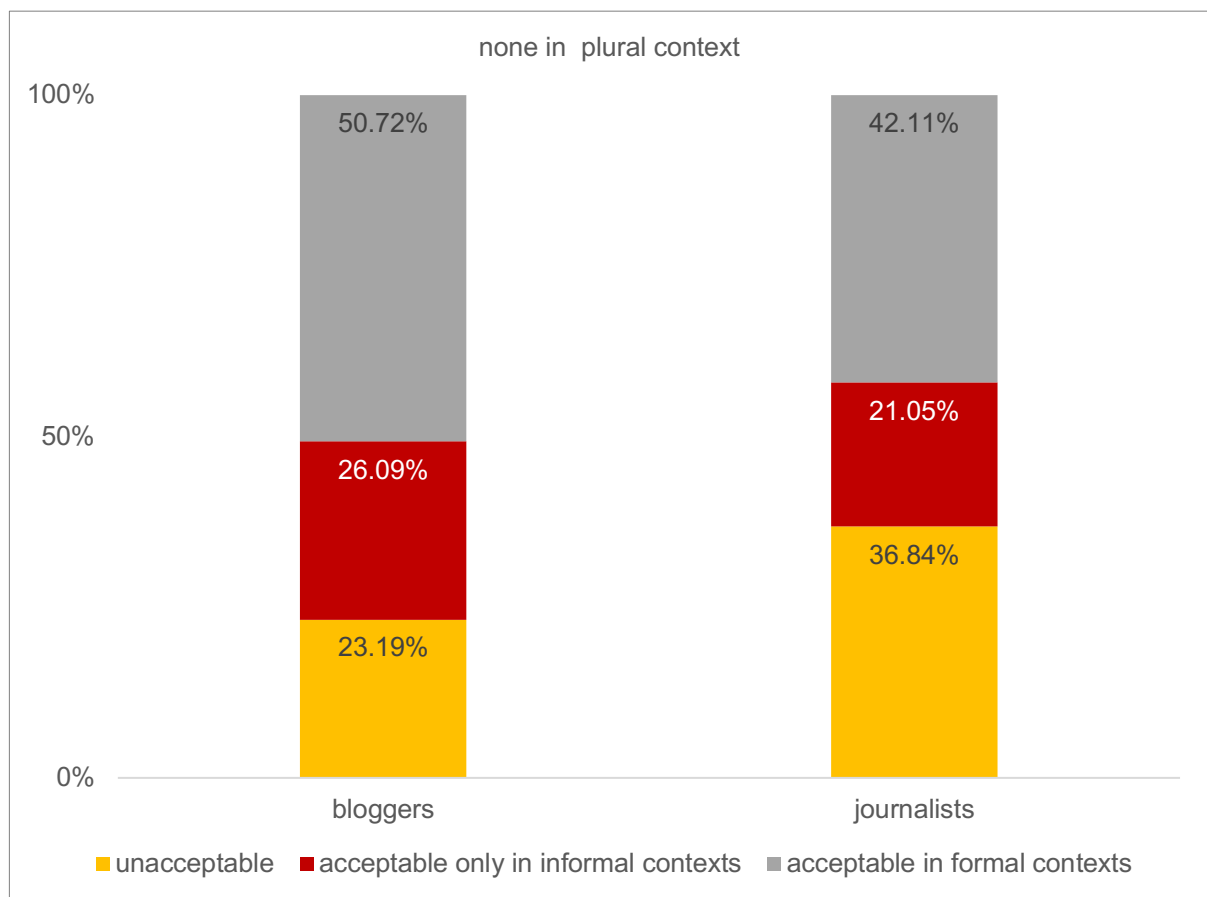


Figure 5.19 Proportion of bloggers and journalists who found *none* used with a plural verb unacceptable, acceptable only in informal contexts, or acceptable in formal contexts

Others noted the need for additional context in order to assess whether the sentence is acceptable or not. One blogger suggested that if the object left on the table were donuts, “it’s fine, because you’re talking about a plural. But if it’s one donut and someone ate the whole

of it, then none was left.” Another blogger called the question ambiguous because a person would need to know “the prepositional phrase that follows ‘none’ to determine [whether it is] singular or plural.”

5.9 Discussion of Results

In the previous sections of this chapter, I have analyzed eight usage problems individually. In this final section of the chapter, I synthesize the findings of the corpus analysis just presented. I focus this synthesis on comparing the usage problems in terms of which are followed proportionally more or less in personal blog writing and news writing, and I discuss which usage problems are viewed more prescriptively by usage guides, bloggers, and news writers. I also discuss the extent to which the attitudes of bloggers and news writers align with the usage patterns observed in this study. Finally, I consider a usage profile for the usage guides in the study, noting which guides tended to treat usage problems leniently and which tended to treat the usage problems more prescriptively.

5.9.1 Comparison of Patterns of Usage

The patterns of prescriptively correct and incorrect usage for the eight usage problems presented in this study showed varying disparities across registers. The SPLIT INFINITIVE usage problem showed the largest disparity in the proportion of instances in which the traditional rule was observed between registers with a difference of 18.00%. The WHO/WHOM usage problem showed the smallest disparity across registers with a difference of only 2.20%. The range of these disparities is 15.80%. On average, the disparity between instances in which the traditional rule was followed in both registers was 8.02% (SD = 5.50%). I use this average as a threshold to objectively distinguish between differences that I consider substantial and insubstantial. That is, any disparity that is equal to or higher than 8.02% I call

substantial; any disparity less than 8.02% I call insubstantial. Of the eight usage problems, register was a substantial factor in three (LAY/LIE, SPLIT INFINITIVE, and NONE IN PLURAL CONTEXT), but not in the remaining five (WHO/WHOM, DIFFERENT TO/THAN/FROM, I FOR ME, SINGULAR THEY, and LESS/FEWER).

When considering the usage problems individually—i.e., considering the average proportion at which each usage problem was followed in both registers—a much higher range can be observed. The WHO/WHOM usage problem showed the highest average of prescriptively correct instances in both registers (97.50%) while the SINGULAR THEY usage problem showed the lowest (18.40%), resulting in a range of 79.10%. On average, the disparity of prescriptively correct usage across usage problems is 69.45% (SD = 30.64%).

The fact that the majority of the usage problems included in this study show an insubstantial difference across registers suggests that blog writing and news writing are generally quite similar in terms of the proportions in which the traditional rules for the eight usage problems under investigation were observed and flouted. These comparisons are visualized in Figure 5.20. Data observed in the blogs corpus appears in solid bars; data observed in the news corpus appears in patterned bars. The solid yellow bars represent the proportion of lines marked correct in the blogs corpus. The solid red bars represent the proportion of lines marked incorrect in the blogs corpus. The patterned lines represent the same information in the news corpus. Below each pair of bars in the figure are two percentages, each in a box. The box on the left, a white, outlined box, shows the average proportion of correct instances observed in both registers. The shaded box on the right shows the percent disparity of prescriptively correct variants between both registers. Values that appear in a light gray box are those for which an insubstantial percent disparity was observed

between registers. Values that appear in a dark gray box are those for which a substantial percent disparity was observed between registers.

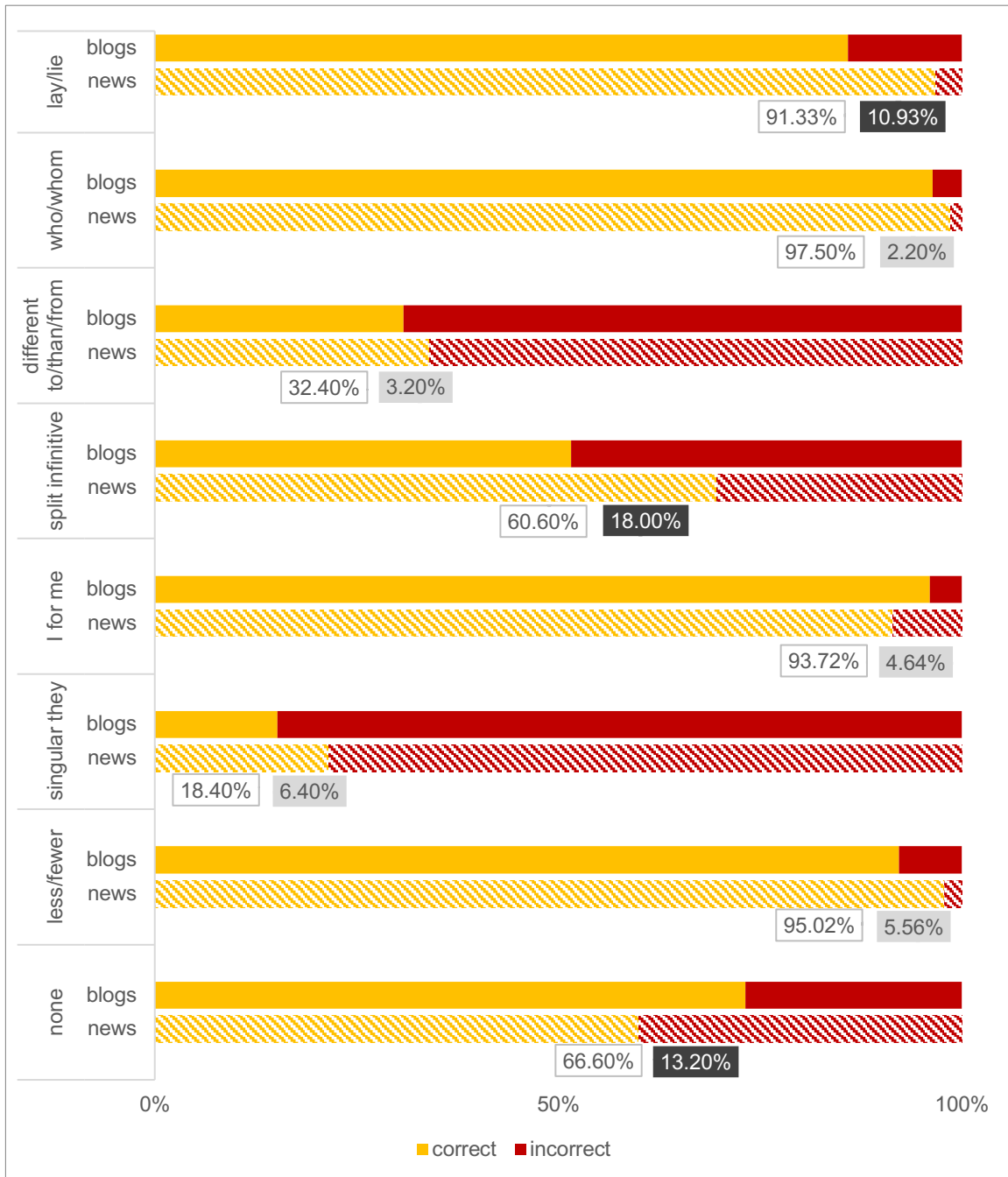


Figure 5.20 Comparison of correct and incorrect instances of eight usage problems in blog writing and news writing. Below each pair of bars, the average proportion of correct instances from both registers appears in an outlined box. The percent disparity of prescriptively correct variants is shown in a shaded box. Values shown in a light gray box are considered insubstantial; values shown in a dark gray box are considered substantial.

Seeing the data in a side-by-side comparison as shown in Figure 5.20 visually demonstrates the higher disparity in the proportion of correctly observed instances across usage problems (shown in the bordered boxes, avg = 69.45%, range = 79.10%) than between registers (shown in the shaded boxes, avg = 8.02%, range = 15.80%). This finding suggests that, in general, the extent to which each usage problem is followed varies when compared with other usage problems, but the formality of the register does not appear to substantially affect the extent to which the rules for these eight usage problems are and are not followed. Put another way, the extent to which the individual usage problems are followed varies to a high degree, but the amount of variation that appears to be the result of register is much less. As Figure 5.20 shows, there was surprisingly little variation in the extent to which the majority of these usage problems were followed in both registers.

Though the proportions for each individual usage problem in both registers are similar, the data in Figure 5.20 shows that the traditional rules of the eight usage problems were flouted slightly more often in blog writing than in news writing in the majority of the usage problems. Only the I FOR ME and NONE IN PLURAL CONTEXT usage problems showed patterns in which the rules were flouted more often in news writing. Possible explanations for this finding for each of these usage problems might be related to the fact that news writing contains a lot of transcriptions of spoken language in which flouting the I FOR ME rule is common, and news writing often treated *none* as a singular pronoun even when it referred to a plural, countable antecedent (which is in line with the conservative prescriptive rule).

Applying the framework shown in Table 4.8 and explained in Section 4.2.3.4, four of the usage problems met the criteria for general rules, two met the criteria for questionable

rules, and two met the criteria for undetermined rules. The usage problems in each category are shown in Table 5.12.

Table 5.12 Classification of usage problems according to formal/informal distinctions

General Rules	Questionable Status	Undetermined
LAY/LIE	DIFFERENT TO/THAN/FROM	SPLIT INFINITIVE
WHO/WHOM	SINGULAR THEY	NONE IN PLURAL CONTEXT
I FOR ME		
LESS/FEWER		

All of the usage problems except for SPLIT INFINITIVE and NONE IN PLURAL CONTEXT fall into the extreme ends of the classification framework. SPLIT INFINITIVE and NONE IN PLURAL CONTEXT fall outside the classification framework. None of the usage problems met the criteria for being classified as “formal” or “leaning formal” and none of the usage problems met the criteria for being classified as “informal” or “leaning informal.” Thus, even though usage guides often make distinctions based on formality—suggesting that a rule may be followed or ignored depending on the formality of the situational context in which the language is produced—such a distinction was not observable in the data used for this study. This finding might suggest that the high level of formality that I argued in Chapter 3 is characteristic of news writing and the high level of informality that I argued in Chapter 3 is characteristic of blog writing is not necessarily reflected in the language of both registers. In addition, this finding could suggest that usage advice based on the formality of the situation is not as helpful as one might assume because it is not always observable in actual language use. Certainly, determining what constitutes formal and informal situations is in large measure subjective, and usage-guide authors likely have differing views on the matter. Therefore, such a finding implies that usage-guide authors might consider reframing their advice in terms of more concrete, defined registers rather than in terms of abstract levels of

formality. Further empirical work would certainly be needed in order for these kinds of recommendations to be possible and useful.

The results of the survey conducted as part of this project can further illuminate the disparity in the assumed level of (in)formality in blog writing and news writing. The survey results demonstrated that while blogs may often be perceived as informal, bloggers themselves do not necessarily view their writing as highly informal. This could be due to a number of factors. The situational characteristics of blogs lean informal in many ways, as I argued in Chapter 3, but many blogs still discuss serious topics. Obviously not all personal blogs are daily diaries or informal discussions of frivolous topics as is stereotypically thought. Table 5.13 shows how respondents to the attitudes survey answered the question of how formal they see their writing. As the data show, journalists generally view their writing as being more formal than bloggers, but roughly half of bloggers (44.64%) still view their blog writing as being at least somewhat formal.

Table 5.13 Perceived level of formality of blog writing and news writing

Level of Formality	Bloggers	Journalists
Very formal	3	2
Somewhat formal	22	18
Somewhat informal	24	6
Very informal	7	0

In addition to the ways that bloggers and journalists perceive the formality of their writing, a majority of both said that they demonstrate a high level of care in composing their writing. In fact a higher proportion of bloggers (74.03%) stated that they carefully edit their posts before publishing them. Only 65.38% percent of journalists said they did so. Certainly the two groups might have differing definitions of what constitutes “careful editing,” and the sample who participated is not representative of all bloggers and all journalists; nevertheless, this

data from the survey suggest that blog writing and news writing may have more in common in terms of formality than originally argued.

5.9.2 Attitudes of Usage-Guide Authors, Bloggers, and News Writers Toward the Usage Problems

Figure 5.21 represents descriptive statistics for each of the usage problems as a box and whisker chart and offers insight into the first research question investigated in this study, which asked about the level of prescriptivism for each usage problem. The “x” inside each box in the figure represents the average prescriptivism rating for each usage problem. The upper and lower boundaries of each box represent the 75th and 25th quartile, respectively. The upper and lower tick marks of the whiskers represent the maximum and minimum values, respectively.

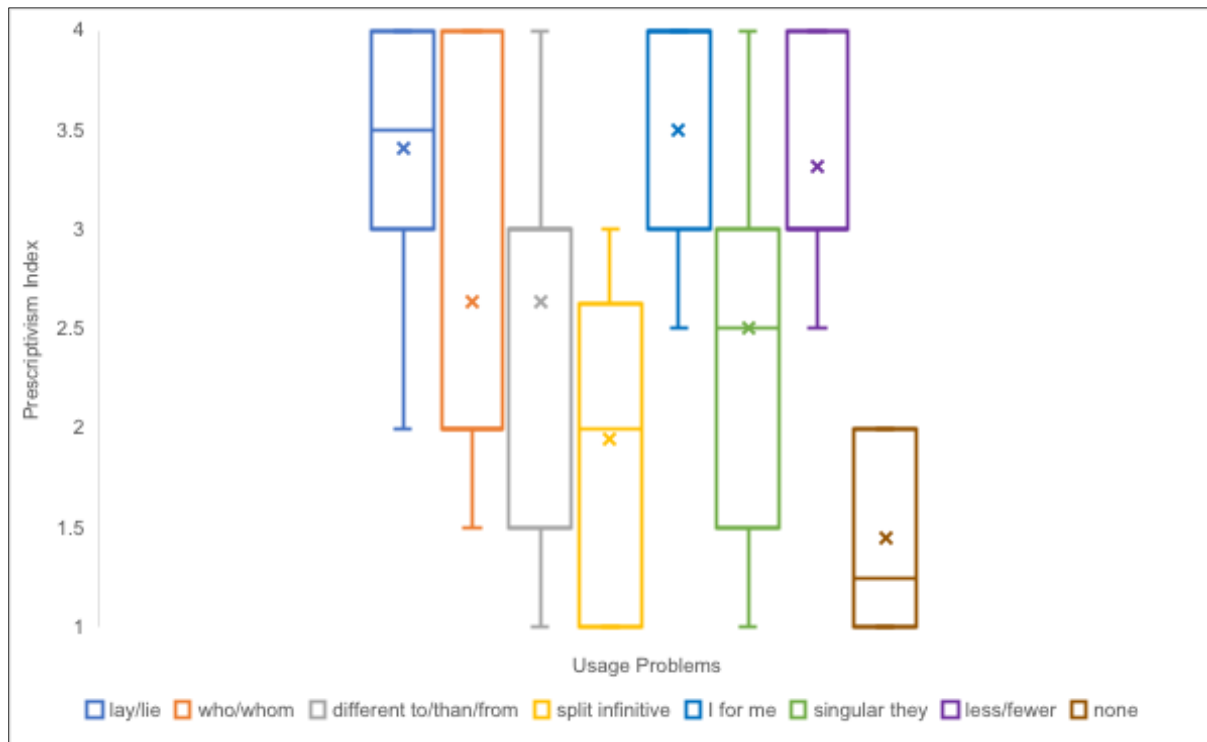


Figure 5.21 Box and whisker plot showing the descriptive statistics for the prescriptivism index of each usage problem

Someone unfamiliar with usage guides might assume that all usage guides simply describe a prescriptive rule and recommend that their readers follow it in all instances. However, the data in Figure 5.21 suggests otherwise. From this data, it is clear that the usage guides offer different advice for different usage problems. Taken on average, the usage guides took the most prescriptive stance for the LAY/LIE, I FOR ME, and LESS/FEWER usage problems with each receiving an average prescriptivism index above 3.00. The NONE IN PLURAL CONTEXT usage problem is treated the least prescriptively among the usage guides with a prescriptivism index of 1.45. In addition, the length of the box for the WHO/WHOM usage problem in Figure 5.21 shows the high amount of variance within the entries that discuss this usage problem. The individual scores for each usage problem from each guide are included in Appendix C.

The prescriptivism profile of the usage guides presented in Figure 5.21 is essentially a profile of the attitudes that usage-guide authors tend to espouse for each usage problem. Figure 5.22 presents the attitude profiles of journalists and bloggers who participated in the survey for this study. A comparison of the two reveals important similarities. For instance, the attitude profile of the survey participants shows the highest levels of prescriptivism for the same usage problems as those that were highest in the usage guide profile, namely, LAY/LIE, I FOR ME, and LESS/FEWER. The solid yellow bars in the figure represent the proportion of instances in which bloggers found flouting a particular rule unacceptable. The solid red and solid gray bars represent the proportion of times bloggers considered a prescriptively incorrect sentence acceptable in different circumstances. The corresponding patterned bars show the same information for journalists who participated in the survey. The longer length of the yellow bars (solid and patterned) for the LAY/LIE, I FOR ME, and

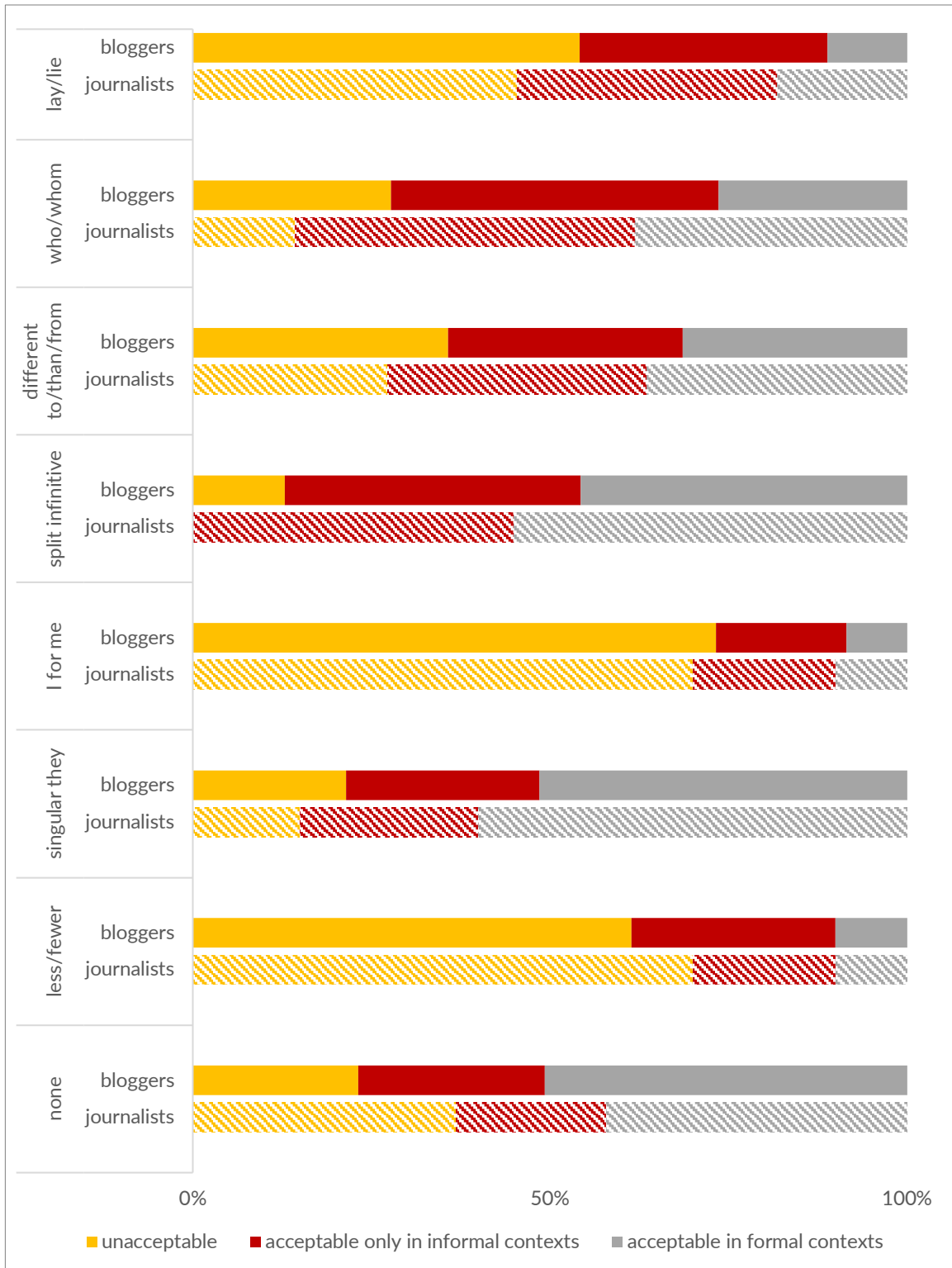


Figure 5.22 Comparison of attitude profiles

LESS/FEWER usage problems show that users generally found flouting the rules for these usage problems to be unacceptable, just as the usage-guide authors did. Whether participants' attitudes are influenced by the usage guides is difficult to say, but it does appear that there is some interaction between the two.

The level of consistency between bloggers' and journalists' attitudes is also noteworthy. In general, the lengths of the colored solid lines are similar to the lengths of the corresponding colored patterned lines, indicating little difference between bloggers' and news writers' attitudes for the usage problems investigated in this study.

While the attitude profiles of both groups were similar, in most cases, the bloggers who participated in the study took a slightly more prescriptive view on the usage problems than the journalists did, as indicated by the solid yellow bars that are slightly longer than the patterned yellow bars in Figure 5.22. The only two usage problems about which journalists took a more prescriptive view than bloggers were LESS/FEWER and NONE IN PLURAL CONTEXT.

5.9.3 Alignment of Usage Patterns with Attitude Profiles

In Section 5.7.4, I noted an interesting discrepancy in the attitudes of bloggers and journalists toward certain usage problems and the usage patterns observed in the corpus analysis. More specifically, I noted that bloggers and journalists generally felt that the rules for some of the usage problems, e.g., WHO/WHOM, SPLIT INFINITIVE, AND NONE IN PLURAL CONTEXT, were generally acceptable to break, but these rules were followed to a proportionally large extent in the corpus data. Comparing Figure 5.22 with Figure 5.20 offers interesting insight into the ways that the usage patterns in the data align with the attitude profiles collected through the survey.

As a comparison of the figures shows, the attitude profiles (Figure 5.22) for the DIFFERENT TO/THAN/FROM and SINGULAR THEY usage problems aligned to a large extent with the usage patterns observed in the corpora (Figure 5.20). That is, the proportions of survey respondents who said that the sentence flouting the rule was unacceptable roughly corresponded to the proportion of times the rule was followed in the corpus data. This correspondence suggests that for these rules, bloggers and journalists consciously perceive that breaking the rule is wrong, so they tend to follow it in their own writing.

For other usage problems, however, the same pattern was not observed in the data. For instance, comparing the data for the SPLIT INFINITIVE usage problem in Figure 5.22 and Figure 5.20 shows a tendency for bloggers and journalists to take a very lenient stance toward this usage problem even though they tended to follow the rule to a comparatively larger extent in their own writing. A similar pattern can be observed for the NONE IN PLURAL CONTEXT usage problem. When asked in the survey why they selected the answer for the level of acceptability, more respondents from both groups said they relied on their ear rather than applying a rule for these usage problems. In other words, a majority of survey respondents did not consciously remember a prescriptive rule governing these usage problems and consider it when determining whether the given sentences were acceptable or not; they simply used their native intuition to determine if the sentences sounded right or wrong. This finding, along with the comparisons of the usage and attitude profiles discussed above, suggests that the prescriptively correct use of SPLIT INFINITIVE and NONE IN PLURAL CONTEXT may come naturally to both groups.

Interestingly, in no circumstances did survey participants demonstrate heavily prescriptive attitudes toward a usage problem (i.e., a high proportion of survey respondents

feeling that flouting the rule was unacceptable) for which a high proportion of writers ignored or broke the rule. Such a pattern, if it were observed, would be indicative of an attitude in which participants were bothered by others' usage, but not necessarily their own. However, only the reverse was observed in this data (i.e., displaying a high degree of leniency toward a rule that a high proportion of writers upheld), as described above.

5.9.4 Profiles of Usage Guides

While not one of the stated goals of this study, an interesting pattern emerged when comparing the level of prescriptivism at the level of the usage guide instead of at the level of the usage problem, as discussed in Section 5.9.2. Table 5.14 shows the average prescriptivism index for each of the usage guides included in this study. As the table makes clear, some usage guides tended to be much more prescriptive in general than others, and vice-versa. Strunk and White (2009) had the highest average prescriptivism index (3.56), indicating that for most of the usage problems analyzed in this study, Strunk and White suggested that their readers should uphold the rule. *MWDEU* (1994) had the lowest average prescriptivism index, suggesting that in most cases, the authors of this usage guide allowed their readers to make their own decisions on whether to follow the traditional rules or not. Knowing approximately how prescriptive or lenient a usage guide is can be useful information for readers, and it can allow them to more critically consider the advice given in the guides, rather than simply accepting the advice that one usage guide may give.

Table 5.14 Average prescriptivism index for each usage guide analyzed for this study

Usage guide	Avg prescriptivism index
<i>Merriam-Webster's Dictionary of English Usage</i> (1994)	1.81
Peters (2004)	2.00
Butterfield (2015)	2.38
Trask (2006)	2.50
<i>The American Heritage Guide to Contemporary Usage and Style</i> (2005)	2.56
Brians (2013)	2.63
Garner (2016)	2.81
O'Connor (2009)	2.88
Fogarty (2008)	3.29
Batko (2004)	3.43
Strunk and White (2009)	3.56

In summary, the results of the current chapter have demonstrated some expected and some surprising results in the usage profiles and attitude profiles of the usage problems under investigation. In the next chapter, I conclude by summarizing the study and addressing implications for technical-editing practice.

CHAPTER 6. CONCLUSIONS AND IMPLICATIONS FOR TECHNICAL EDITING

In this project, I have offered an empirical analysis of eight usage problems and the extent to which their use differs in two registers of writing—one that I have argued is representative of formal writing and one that is representative of informal writing. In this chapter, I present a final discussion of the results from these eight analyses organized according to the research questions presented in Section 3.4. As part of this discussion I outline key limitations of this study and offer ideas for further research to fill the gaps left by these limitations. I conclude the chapter with a discussion of the implications that empirical research on prescriptive usage problems can have for technical editing.

6.1 Review of Research Questions

6.1.1 RQ1

The first research question in this study investigated the extent to which a set of 11 American usage guides recommend following the traditional rules for eight well-known usage problems. To answer this question, I developed a four-point scale to code the prescriptivism index for entries dealing with each usage problem from 11 current usage guides. A trained rater and I then coded each entry, after which I calculated a prescriptivism index for each usage problem in each guide by calculating the average of the two ratings.

The findings from this analysis revealed a considerable amount of variation across usage problems (see Section 5.9.2). It also revealed some similarity in the level of prescriptivism within usage guides (see Section 5.9.4). Usage guides were ranked according to the average level of prescriptivism for the entries we coded, with some guides taking a generally more prescriptive approach than others.

While this analysis for RQ1 has been robust in many ways, it is not free from limitations. For example, in this study, I collected entries from only a small number of current usage guides in order to assess current recommendations for these usage problems. Future studies might collect entries from many more usage guides to gain a more comprehensive profile of the level of prescriptivism that can be observed in usage guides for these eight usage problems. In addition, in the current study, raters assigned only one prescriptivism index for each usage problem per usage guide (see Section 4.1.5). Future studies might identify individual comments in the usage guides (see Yañéz-Bouza, 2015), assign a rating to each comment in which a usage problem is discussed, and then average these ratings together in order to create a potentially more accurate prescriptivism index for each usage guide.

6.1.2 RQ2

The second research question undertaken in this study investigated the degree to which formal edited writing (news) and informal unedited writing (personal blogs) differ in the ways they conform to the prescriptive usage advice given for the same set of eight usage problems. To answer this question, I extracted KWIC lines that contained the possible variants related to each usage problem from a corpus of news writing (COCA–N) and a corpus of personal blog posts (CPBP) compiled for the purposes of this study. I then coded a near-random sample of the KWIC lines associated with each usage problem for different factors that allowed me to determine if each instance upheld the prescriptive rule or flouted it. Instances that upheld the rule were marked as “correct” while those that flouted the rule were marked as “incorrect.” (The instances coded for the NONE IN PLURAL CONTEXT usage problem were coded as matching or not matching as described in section 5.2.8.) I then

compared the proportion of correct and incorrect instances of each of the eight usage problems included in this study. The results of this analysis showed little variation across registers in the usage patterns observed for the eight usage problems under investigation.

In general, the proportions of prescriptively correct and incorrect instances were similar across registers for each usage problem. There was, however, considerable variation in the extent to which individual usage rules were followed. Some usage rules were flouted to a proportionally high degree (e.g., SINGULAR THEY) while others were observed to a proportionally high degree (e.g., WHO/WHOM). However, as previously noted, usage rules that were flouted to a high degree were flouted in both blog writing and news writing to a relatively high degree, and usage rules that were observed to a high degree were observed to a relatively high degree in both registers. This finding, along with the fact that only three of the eight usage problems showed rates of variation between registers higher than the average (see Section 5.9.1) suggests that register is not a considerable factor for observing variation in these usage problems.

The data analyzed to study each usage problems consisted of a relatively small number of KWIC lines (250 instances of each variant in most cases). Coding a larger sample of variants would provide a more complete and accurate profile of the patterns of use of each usage problem.

6.1.3 RQ3

The final research question in this study compared the ways in which blog writers and news writers perceive the usage problems investigated in this study. From the analyses presented in Chapter 5, there is in general little variation between bloggers' and journalists' views with respect to the usage problems they were asked about, though in the variation that

was observed, it was often the case that bloggers took a more prescriptive approach than journalists. This surprising finding maybe due in part to the fact that the journalists who responded to the survey were slightly younger on average than the bloggers who responded. Because youth is generally associated with linguistic innovation (see Tagliamonte, 2016), younger people are likely more accepting of language change. Or it could be that journalists, because they belong to a professionalized community and because they are trained writers, are able to draw from their position and training to build their ethos. As the saying goes, a person must first know the rules before they are allowed to break them. Journalists' status as knowers of the rules may have given them the privilege to be more lenient when considering the extent to which the rules should be followed. Bloggers—particularly those who blog about current events or other more serious topics—do not have the advantage of the instant credibility that comes with writing for an established news organization. In this way, bloggers may need to rely more heavily on correct usage in order to build their credibility.

The original goal of this project was to compare formal and informal writing. In Chapter 3 of this dissertation, I argued that the situational characteristics of news writing contribute its formal nature and that the situational characteristics of personal blog writing contribute to its informal nature. While that argument still holds true to some extent, the analysis presented in Chapter 5 has revealed much fewer linguistic differences than anticipated based on these differing levels of formality. This lack of difference is likely due to what seems to be a hybrid nature in both of the registers. On the one hand, blog writing—even personal blog writing—can often take on traditionally more formal situational characteristics, such as discussing religious beliefs, explaining technical information, or reporting on current events. Indeed, nearly one-fifth (19.36%) of the blog posts that were

coded for their situational characteristics (see Section 4.2.2.4) were coded by both coders as discussing one of these three topics. On the other hand, news writing often contains quoted spoken material, which can include casualisms, slips of the tongue, and other forms of colloquial or informal language that are commonly found in spoken discourse. As noted in the analysis presented in Chapter 5, journalists are trained to include quotations exactly as they were spoken, even if that means leaving in mistakes related to grammar or usage. The large amount of quoted material contained in news writing is one major factor that decreases the formal nature of news writing because it introduces a high level of informality into the discourse. Unfortunately, it was not possible to account for quoted material in the current project. The nature of the data from the news corpus made it impossible to reliably identify and remove quoted material in an automatic way, and time constraints precluded manual removal of quoted material. However, future studies might use a much smaller corpus of news writing with all quoted material clearly annotated and structured in a way that would make its removal using automatic methods possible. Follow-up studies investigating the same usage problems presented here would provide a clearer picture of whether the language produced by news writers—the formal language that is planned, revised, and edited—contains roughly the same proportion of errors as reported in this study or if the proportion of errors observed is different. Observing a much lower proportion of errors in such a study would provide evidence that the errors observed in news writing as a whole are largely the result of transcribed spoken language in the form of interview responses.

Another follow-up study might investigate the writings of journalists in different registers. For instance, some journalists maintain blogs that are not affiliated with their news organization. Others write books or columns in magazine articles. Many maintain a presence

on social media—particularly Twitter—in which they share updates about stories they are reporting. All of these different registers have different situational characteristics that affect the language produced. A fully complete understanding of the level of prescriptivism that can be observed in news writing would need to include the results from studies that investigate the writings of journalists in these other registers. Such an area of study is ripe for further investigation.

Taken together, these findings offer valuable insight into the ways that prescriptive usage rules are and are not followed in formal and informal writing. To conclude this project, I discuss some of the implications that empirical research of the type presented in this study can have for disciplines related to writing studies with particular attention paid to technical communication and technical editing. My purpose in doing so is to argue that studies that use empirical methods—such as corpus linguistics—to study prescriptivism in actual language use across registers can encourage more complete rhetorical decision making among technical communicators and editors.

6.2 Implications for Technical Editing

6.2.1 Favoring a Rhetorical Approach to Technical Editing

In her article on the ways that technical communication textbooks fail engineering students, Wolfe (2009) summarized some of Tufte's (1983) foundational maxims of data design. About these rules, Wolfe says, "*Although we do not want to teach our students blind adherence to these maxims*, students can nonetheless benefit from learning that such principles carry strong currency among certain audiences" (p. 364, emphasis mine). Though Wolfe specifically referenced data-design maxims in her comment, I believe her argument can be usefully extended to all principles of technical communication, including prescriptive

language rules and issues of contested usage. Technical communication is, after all, a humanistic discipline (see Miller, 1979), one in which blind adherence to *any* rule or maxim should be regarded as suspect. Instead of knowing only how to mindlessly adhere to a set of rules, technical communicators—including technical editors—should know the rules to an extent that allows them to make rhetorically informed choices about, as Wolfe said, the kind of currency they carry with certain audiences. Such an understanding of different rules, I argue, requires empirically investigating these rules and the contexts in which they are discussed, followed, and challenged.

These opposing views (i.e., adhering blindly to a set of rules versus making rhetorically informed choices about whether or not to follow a set of rules), are related to two rhetorical concepts: correctness and decorum. Correctness and decorum are two of the five virtues of style in rhetorical theory. On the surface, they appear to have conflicting goals. Correctness is concerned with “adher[ing] to the conventions of vocabulary and syntax, grammar and usage, that predominate in [a language]” (correctness, n.d.). Decorum, on the other hand, is concerned with propriety or appropriateness. In most instances of language use, the correct choice is also the decorous choice because the linguistic features in question are not contested in terms of their correctness. For example, if an editor notices an instance of *your* in a writer’s article where a sense of *you are* is intended, the editor’s choice to change the word from *your* to *you’re* is both decorous and correct because there is no serious, large-scale discussion currently taking place in which people are arguing that *your* should be an acceptable alternate for *you’re*. However, in other instances of language use—those in which the two alternates are more heavily contested in terms of correctness—these two virtues are in conflict, requiring editors to decide which one to favor. For example, if an editor notices

an instance of *they* used as a singular pronoun in a sentence, they must choose to favor correctness by either changing the antecedent to its plural form or replacing the pronoun with a singular form (which some usage manuals suggest), or they must choose to favor decorum and let the singular *they* stand.

It is a common stereotype for editors to be seen with an attitude that favors correctness over decorum regardless of context. Indeed, Rude and Eaton (2011) identified one of the four primary responsibilities of copy editors as making a document correct. In this stereotypical view, editors memorize a set of rules contained in a style guide and blindly correct a manuscript, marking any features that deviate from the rules they have learned. A perhaps less common but certainly more useful perception of technical editors is one in which they recognize the importance of decorum and allow this principle to influence the decisions they make—whether to enforce a rule contained in a style guide or not. Buehler (2003) distinguished between a “rhetorical” and “programmatic” approach to editing, advocating that technical editors should prefer the rhetorical approach. In Buehler’s rhetorical model, the editor makes editorial decisions based on the rhetorical situation of the communicative event while in the programmatic model, the editor simply “applies a set of rules to all situations...without concern for all the varied elements of the situation itself” (p. 459).

Other scholars have called for a similar emphasis on the rhetorical underpinnings of technical communication and language use. Miller (1979) argued that technical communication “should present mechanical rules and skills against a broader understanding of why and how to adjust or violate the rules.” (p. 617). Earlier, Gorrell (1977) argued that “usage is not a question of grammar or linguistics or morality, but of rhetoric” (p. 20). He

continued: “[U]sage labels or other facts about language...contribute useful knowledge to any rhetorical choice, but do not automatically make the choice” (p. 23). For technical editors, then, considering the rhetorical situation allows them to confidently make choices about what is appropriate or decorous, which, as noted above, may not always be considered correct. In other words, the prescriptions outlined in usage guides and style manuals should not be taken as inflexible rules, never to be disobeyed. It is still up to writers and editors to consider the rhetorical situation within which the communicative event takes place so that they can then make the choices that they feel will best meet a given exigence for a given audience within given constraints (see Bitzer, 1968).

Of course, in order for technical editors to make a decorous or rhetorical choice, they must have information about exigences, audience(s), and constraints of the situation in order to accurately gauge what will be considered appropriate. These elements all relate to the rhetorical situation as Bitzer (1968) defined it. I argue that one additional element that can help technical editors make decorous choices in their work is empirical data.

6.2.2 Using Empirical Data to Inform a Rhetorical Approach to Technical Editing

Empirical research on prescriptive usage problems—such as that presented in this study—encourages a more critical examination of usage rules that technical editors encounter regularly in their work. Empirical research on this topic can help technical editors make more informed choices when considering the language they allow in the technical documents they edit by helping them understand that not all prescriptive language rules contained in style manuals need to be observed in all cases. Empirical research can help pinpoint which cases these might be.

Some current usage guides have begun to incorporate empirical data from corpora in their work. Peters (2004), Garner (2016), and Butterfield (2015) all use corpora to some extent in order to gather data about how contested usage items can be observed in actual language use. However, additional data is needed in order to more fully describe if and how usage differs across registers. The current study aims to contribute to this type of description.

In many instances, the additional empirical information provided in usage guides and style manuals is helpful and informative. In other instances, however, it is presented in a way that ignores what the data might suggest. For example, in *The Chicago Guide to Grammar, Usage, and Punctuation*—a book Warren (2017) “highly recommended” (p. 263) in his *Technical Communication* review—Garner (2016) devoted a section to word usage in which he provides guidelines for troublesome words, sometimes accompanied by usage data from Google ngrams.¹⁷ In this way, Garner showed empirical support for many of his suggestions, but he did so in a way that does not always encourage readers to take a critical view of these usage problems. The ngram chart that accompanies his entry on *home in*, for instance, shows that the prescriptively correct *home in* is declining in frequency, while the disparaged variant *hone in* has been increasing over the years. Yet Garner made no mention of this convergence, instead simply calling the preferred phrase “frequently misrendered” (p. 280). Merriam-Webster calls *hone in* “established” in American English (“hone in”, n.d.), so it is entirely possible that for some audiences, reading *home in*, as Garner advocates with no caveats, would seem an error and draw unnecessary attention to itself, a rhetorical result few would set out to achieve.

¹⁷ <https://books.google.com/ngrams>

Usage problems and issues of prescriptivism in general extend beyond issues at the language level such as clarity, precision, or accuracy. They also reflect and sometimes promote social problems as well, making it even more critical that technical editors take a rhetorical approach in their understanding of prescriptive language rules. Ebner (2017) emphasized the social nature of usage problems and the divisive role that usage problems carry out in a society, and Chrisomalis (2015) stated that general prescriptivism “is used to index social status and moral propriety through appeals to tradition and authority within a national or even international linguistic context” (p. 68). Many usage problems are long-standing shibboleths used by groups in power to suppress people in populations who acquired a variety of English different from theirs. Maintaining and perpetuating these shibboleths has social consequences, and it is important for technical editors to be aware of these concepts so they can consciously determine how they might work to change them. Of course, it is not appropriate for technical editors to disregard all conventions of Standard English as a display of social protest, but it is important for technical editors to be able to understand and critically consider the language rules they enforce, including (and maybe even especially) the usage problems they encounter often in their work.

Many technical editors likely do not feel that they have the authority to violate usage rules for rhetorical purposes. However, as Graves and Graves (1998) argued, technical communicators play an important ethical role in ensuring that the language used in technical documents does not promote offensive or harmful social problems, such as racism, classism, or sexism—even when such language falls in line with approved guidelines. According to Graves and Graves,

Editors, writers, and researchers of technical communication are in a unique position to use their expertise not only to produce top quality documents, but also to examine and raise for discussion those linguistic constructions and conventions that portray reality in questionable ways. (p. 412)

Because prescriptive language rules are not learned through natural processes of language acquisition but instead through conscious pedagogical instruction (Bourdieu, 1991, p. 61), they covertly operate within a frame of classism, helping to sustain existing structures of social power by benefiting those who grow up in environments where parents teach these rules to their children or to people who are able to attend the best schools. Some usage problems are also overtly linked to problematic and harmful ideologies. For example, SINGULAR THEY, as an alternative for the supposedly gender-neutral *he* and a pronoun for people who do not identify as male or female, is widely discussed within the framework of sexism and LGBTQ+ rights. Technical writers must understand these issues and consider them in their work, and technical-writing teachers must help their students understand that technical writing does not reveal absolute reality (Miller, 1979, p. 616), but instead functions rhetorically and therefore requires critical thinking, audience awareness, and conscious decision-making when considering whether to follow or flout prescriptive usage advice (see Buehler, 2003; Connaster, 2004).

6.2.3 Describing the Benefits of a Rhetorical Approach to Technical Editing

Thus far, I have argued that technical editors should favor a rhetorical approach to their work—one that allows them to make decisions that may emphasize the virtue of decorum over the virtue of correctness, and I have discussed some of the ways in which empirical data can complement a rhetorical approach to technical editing. What benefits

might technical editors and the field of technical communication in general enjoy by making use of empirical studies of usage problems and adopting a rhetorical approach to technical editing? Rigorous empirical work—the kind that goes beyond simple ngram searches like those Garner (2016) conducted as discussed above—offers data not only about the frequency at which usage problems are or are not followed but also about the situations in which these patterns can be observed. This kind of information offers deeper insights into the ways that different usage problems are treated in actual language use, which will help practitioners make more informed rhetorical decisions about their own language use—in spite of what the rules might say. Additionally, empirical work and a rhetorical approach can be helpful for authors of popular technical-communication textbooks and general style guides as they revise their works for future editions. As these works become more linguistically and socially informed, the teachers who use them in their classrooms are likely to, over time, adopt this view and share it with their students.

In order for this change to take place, more writing handbooks and usage guides should base their advice on current language use, rather than the impressions or personal peeves of the authors, or the recycled advice of previous guides. This argument is, of course, not new. Meyers (1995) cited this topic playing prominently at meetings of the Conference on College Composition and Communication as early as 1935. Through his own analysis of 40 years' worth of writing manuals, Meyers found that little had changed in the ways that writing handbooks discussed prescriptive usage rules, and he called for revision that reflected the ways these rules are actually used so as to avoid “the ultimate absurdity: usage without users” (p. 62).

One of the benefits of using corpus methods to study actual usage is the systematic way it allows researchers to study actual language use on a large scale. And taking a register perspective to corpus studies allows scholars to interpret their findings within the rhetorical situation that the communicative event took place. In an early study on usage problems, Meyers (1972) used the Brown corpus to study a set of 29 usage problems commonly discussed in writing handbooks. His purpose in doing so was to “find out...what professional, modern, edited, American usage is in some of the problem areas” (pp. 155–156). Though not stated explicitly, a secondary purpose of this study is arguably to provide evidence that handbook writers may find useful in composing or revising their handbooks. If some of the usage problems are seldom followed in edited usage, why continue prescribing them?

In contrast to Meyers’s (1972) study, Dant (2012) carried out an empirical study of usage prescriptions in the *Chicago Manual of Style* with the explicitly stated purpose of making recommendations that may help to shape future editions of the manual. Based on her comparison of 85 of *Chicago*’s prescriptions with actual usage as recorded in the Corpus of Contemporary American English (Davies, 2008), Dant made five recommendations:

1. Retain prescriptions that were observed to be followed at least 80% of the time across all five registers represented in COCA (spoken, news, magazine, academic, fiction). Examples from Dant’s analysis include favoring *accompany by* instead of *accompany with* and preferring *before* or *until* in place of *prior to*.
2. Acknowledge differences among registers when a prescription is followed less than 80% of the time in one or more registers. E.g., favoring *different from* over

different than, which was followed 53% of the time in spoken language and 74% of the time in news writing, but more than 80% in all other registers.

3. Refine prescriptions when the prescription does not reflect actual usage, when no mainstream variant can be observed, or the disfavored variant seems to have eclipsed the favored one. E.g., *Chicago* prefers *sneaked* as the past tense of *sneak*, yet Dant found that both forms are used in almost split distribution in fiction and magazine writing and *snuck* is a “vigorous variant” (p. 35) in newspaper and academic writing. In spoken language, *snuck* is more common, though *sneaked* is still a vigorous variant.
4. Drop prescriptions that are flouted at least 40% of the time in all written registers. E.g., *Chicago*’s proscription of *anxious to* in favor of *anxious about* is not observed in actual usage. In all registers, Dant found that *anxious to* occurs at least 74% of the time, which she argues should qualify *anxious to* for “accepted status, or, minimally, the prohibition against it should be dropped” (p. 37).
5. Drop prescriptions that are followed 100% of the time in all five registers. E.g., *inhere* colligates with the prescribed particle *in* (not the proscribed particle *within*) in all instances found in COCA. Prescriptions like this one, Dant argues, should be dropped to make room for discussions on “usage issues that are in earlier stages” (p. 37).

Dant’s work summarized here is one example of an empirical study that shows how critically examining usage prescriptions by comparing them to actual language use can help writers, writing teachers, and even the authors of the usage manuals and handbooks that writers rely on. Her inclusion of explicit recommendations offers a clear path that other researchers can

follow, even if they do not adopt the same specific thresholds and recommendations she does. For example, future researchers may agree with Dant's recommendation to drop prescriptions that are observed to be followed 100% of the time but disagree that prescriptions for rules that are flouted at least 40% of the time should be dropped as well. In my estimation, dropping prescriptions that are observed to be flouted less than half of the time would do a disservice to editors who seek information on a particular usage for which they often observe variation. In these cases, it may be better to retain the entry, but refine the actual prescription—a course that Dant may favor, though it is unclear, since her suggestion to “drop” certain prescriptions could mean that the guide should no longer address the issue or that it should still include an entry for the usage problem but simply adjust its recommendation. Future research should clearly define which of these alternatives is recommended.

Considering the results of the present research as well as research like Dant's (2012) that has come before, I conclude by offering what I consider to be three important elements of an ideal usage guide:

1. **All recommendations in an ideal guide should be based on actual usage data, not on the opinions or personal peeves of the usage-guide author(s).** Existing usage guides (e.g., Peters, 2004; Garner, 2016) make use of usage data in determining the advice they offer. However, neither Peters nor Garner considers empirical evidence for every recommendation made (see, e.g., the entry for BETWEEN OR AMONG in Peters or the entry for HYPERTENSION in Garner). Of course, the amount of time it would take to carry out an in-depth corpus-based study for every usage problem discussed in a comprehensive usage guide like

Peters and Garner would be prohibitive, and both Peters and Garner deserve credit for the amount of empirical evidence they do consider in their respective usage guides. A more useful approach for some audiences, however, would be to focus the scope of a given usage guide to include only usage problems that are seen as especially important for writing in a given register or small number of related registers (see Recommendation 2 below) and carry out the necessary empirical research to make data-driven recommendations for each entry.

2. **Usage guides should be written for specific registers.** General usage guides, i.e., usage guides that are intended for general purposes, are certainly useful. Indeed, many of the most popular guides currently in use have a general or comprehensive scope. However, in order to make empirically informed recommendations for each entry feasible (see Recommendation 1 above), a guide should limit its scope to include recommendations for one or more related registers. Even though the current study found that effect of register was insubstantial for most of the usage problems included in this study (see Section 5.9.1), such may not be the case for other registers that might be investigated in future studies. Even considering the current study, the differences between registers should not be completely discounted. Therefore, an ideal usage guide should be targeted for writing in register(s) that are clearly defined by their situational characteristics, similar to the ways the registers under investigation in the present study are defined in Chapter 3, and not vaguely referred to as simply “formal” or “informal.” As discussed in Section 6.1.3, labeling a given register as entirely formal or informal is difficult if not impossible. The more focused

register approach recommended here would decrease the reach of the ideal usage guide but would make it more useful to people who often write documents or messages in specific registers. For example, the data from the current study could be used to write a usage guide specifically for people who write news articles and blog posts. News-specific guides and style manuals exist (e.g., *AP Stylebook* and the *New York Times Manual of Style and Usage*), but they seldom if ever offer usage-based evidence for their recommendations. Those guides that do offer usage-based evidence for their recommendations often are not tailored to a specific register. As a result, there is still a need for usage guides with empirically informed recommendations that are targeted to certain specially defined registers.

3. **Usage-guide entries should avoid biased or evaluative language.** Many popular usage guides contain highly evaluative—often quite derogatory—language to describe those who fail to adhere to the prescriptions set forth in the guide. It may very well be the case that biased and evaluative language appeals to a sinister side of those who read the guides, creating in them a feeling of linguistic superiority. This type of language may sell usage guides, but it also has the damaging effect of promulgating harmful ideologies about varieties of English that are different from the currently accepted standard. As a result, an ideal usage guide should avoid this kind of language. And because the recommendations in the hypothetical ideal guide are based on empirical data drawn from actual usage in specifically defined contexts, the need for justifying recommendations using derogatory terms based solely in opinion would be eliminated.

Producing a usage guide that adheres to the criteria described above will require much additional work. While studies like Meyers's (1972), Dant's (2012), and the project presented in this dissertation offer a start, there is still much more that needs to be done to build a fuller understanding of the ways that actual language use differs from prescriptive usage advice, and in what rhetorical situations or registers. Further empirical studies might begin filling this gap by observing, for example, how common usage problems can be observed across different genres of technical writing. The results of such studies would provide useful data that may eventually have some influence on the way that technical writing is taught in schools and the ways that technical writers and editors do their work.

REFERENCES

- A Dictionary of Modern English Usage. (n.d.). In *Wikipedia*. Retrieved July 25, 2019, from https://en.wikipedia.org/wiki/A_Dictionary_of_Modern_English_Usage
- The American Heritage guide to contemporary usage and style*. (2005). Boston: Houghton Mifflin.
- Anderwald, L. (2016). *Language between description and prescription: Verbs and verb categories in nineteenth-century grammars of English*. Oxford University Press.
- Anthony, L. (n.d.). ClawsAnt (Version 2.0.0) [Computer Software]. Tokyo, Japan: Waseda University. This version of the software is not publicly available; I gratefully acknowledge the developer's generosity in sharing it with me.
- Anthony, L. (2014). AntConc (Version 3.4.3m) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/>
- Auer, A., & González-Díaz, V. (2005). Eighteenth-century prescriptivism in English: A re-evaluation of its effects on actual language usage. *Multilingua*, 24, 317-341.
- Balhorn, M. (2009). The epicene pronoun in contemporary newspaper prose. *American Speech*, 84(4), 391-413.
- Baltimore Sun. (2018). *Baltimore Sun* newsroom directory [Organization chart]. Retrieved from <http://www.baltimoresun.com/about/contact/bal-baltimore-sun-newsroom-directory-ngux-htmlstory.html>
- Batko, A. (2004). *When bad grammar happens to good people: How to avoid common errors in English*. Franklin Lakes, NJ: Career Press.
- Baugh, J. (2001). A dissection of style-shifting. In P. Eckert and J. R. Rickford (Eds.) *Style and sociolinguistic variation*. (pp. 109-118). Cambridge University Press.
- Beede, P., & Mulnix, M. W. (2017). Grammar, spelling error rates persist in digital news. *Newspaper Research Journal*, 38(3), 316-327.
- Bell, A. (1991). *The language of news media*. Cambridge, MA: Blackwell.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
- Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.
- Biber, D. & Egbert, J. (2018). *Register variation online*. Cambridge University Press.

- Biber, D., & Finegan, E. (1994). Introduction. In D. Biber and E. Finegan (Eds.) *Sociolinguistic perspectives on register*. (pp. 315–347.) New York: Oxford University Press.
- Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge University Press.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- Biber, D., Egbert, J., & Davies, M. (2015). Exploring the composition of the searchable web: A corpus-based taxonomy of web registers. *Corpora* 10(1), 11–45.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English*. Harlow, England: Pearson Education.
- Bitzer, L. (1968). The rhetorical situation. *Philosophy and Rhetoric*, 1(1); 1–14.
- Bloomfield, M. (1953). The problem of fact and value in the teaching of English. *College English*, 15(1), 33–37.
- Boettger, R. K., & Wulff, S. (2014). The naked truth about the naked this: Investigating grammatical prescriptivism in technical communication. *Technical Communication Quarterly*, 23(2), 115–140.
- Bourdieu, P. (1991). *Language and symbolic power*. Cambridge: Polity Press.
- Brians, P. (2013). *Common errors in English usage (3rd ed.)*. Portland, OR: William, James & Co.
- Brown, P., & Fraser, C. (1979). Speech as a marker of situation. In K. R. Scherer and H. Giles (Eds.) *Social markers in speech*. (pp. 33–62). Cambridge University Press.
- Buehler, M. F. (2003). Situational editing: a rhetorical approach for the technical editor. *Technical Communication*, 50(4), 458–464.
- Butterfield, J. (Ed.). (2015). *Fowler's dictionary of modern English usage*. Oxford University Press.
- Cameron, D. (1995). *Verbal hygiene*. New York: Routledge.
- Cameron, D. (1996). When words collide: Expert and popular discourse on language. *Language Sciences*, 19(1), 7–13.
- Chapman, D. (2010). Bad ideas in the history of English usage. In R. A. Cloutier, A. M. Hamilton-Brehm, W. A. Kretzschmar, Jr. (Eds.) *Studies in the history of the English language V: Variation and change in English grammar and lexicon: Contemporary approaches* (pp. 141–155). Berlin: De Gruyter Mouton.

- Chrisomalis, S. (2015). What's so improper about fractions? Prescriptivism and language socialization at Math Corps. *Language and Society*, 44, 63–85.
- Christian, D., Jacobsen, S., & Minthorn, D. (2010). *The Associated Press stylebook and briefing on media law*. New York: The Associated Press.
- Connatser, B. R. (2004). Reconsidering some prescriptive rules of grammar and composition. *Technical Communication*, 51(2), 264–275.
- correctness. (n.d.). "Silva Rhetoricae". Retrieved from rhetoric.byu.edu.
- Crystal, D. (1995). *The Cambridge encyclopedia of the English language*. Cambridge University Press.
- Crystal, D. (2006). *Language and the internet* (2nd ed.). Cambridge University Press.
- Curzan, A. (2002). Teaching the politics of Standard English. *Journal of English Linguistics*, 30(4), 339–352.
- Curzan, A. (2014). *Fixing English: Prescriptivism and language history*. Cambridge University Press.
- Dant, D. R. (2012). Using COCA to evaluate *The Chicago Manual of Style's* usage prescriptions. In J. Mukherjee & M. Huber (Eds.) *Corpus linguistics and variation in English: Theory and description*. (pp. 29–39). Amsterdam: Rodopi.
- Davies, M. (n.d.). Introducing the 1.9 billion word Global Web-Based English Corpus (GloWbE) [Blog post]. Retrieved from <https://21centurytext.wordpress.com/introducing-the-1-9-billion-word-global-web-based-english-corpus-glowbe/>
- Davies, M. (2008-) *The Corpus of Contemporary American English (COCA): 520 million words, 1990–present*. Available online at <https://corpus.byu.edu/coca/>.
- Davies, M. (2011) N-grams data from the Corpus of Contemporary American English (COCA). Downloaded from <http://www.ngrams.info> on February 6, 2018.
- Davies, M. (2013). *Corpus of Global Web-Based English: 1.9 billion words from speakers in 20 countries (GloWbE)*. Available online at <https://corpus.byu.edu/glowbe/>.
- Dichtel, F. (2016). A quantifier used on many occasions: *Many* evoking diversity in positive sentences. *International Journal of Corpus Linguistics*, 21(1), 80–104.
- Ebner, C. (2017). *Proper English usage: A sociolinguistic investigation of attitudes towards usage problems in British English* (Doctoral dissertation). Retrieved from LOT Netherlands Graduate School of Linguistics (LOT No. 468).

- Egbert, J., Biber, D., & Davies, M. (2015). Developing a bottom-up, user-based method of web register classification. *Journal of the Association for Information Science and Technology*, 66(9), 1817–1831.
- Faber, B. (2002). Professional identities: What is professional about professional communication? *Journal of Business and Technical Communication*, 16(3), 306–337.
- Fernández-Pena, Y. (2017). 'These points stated, a number of problems remain': A corpus-based analysis of the idiomatisation of collective noun-based constructions. *Exploring Recent Diachrony: Corpus Studies of Lexicogrammar and Language Practices in Late Modern English (Studies in Variation, Contacts and Change in English, 18)*. Edited by S. Hoffmann, A. Sand & S. Arndt-Lappe. Helsinki: VARIENG. Retrieved from <http://www.helsinki.fi/varieng/series/volumes/18/fernandez-pena/>
- Finegan, E. & Biber, D. (1994). Register and social dialect variation: An integrated approach. In D. Biber and E. Finegan (Eds.) *Sociolinguistic perspectives on register*. (pp. 315–347.) New York: Oxford University Press.
- Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions*. Hoboken, NJ: John Wiley & Sons.
- Fogarty, M. (2008). *Grammar Girl's quick and dirty tips for better writing*. New York: Henry Holt and Company
- Follett, W. (1966). *Modern American usage: A guide*. New York: Hill & Wang.
- Garner, B. A. (2016). *Garner's modern English usage* (4th ed.). Oxford University Press.
- Garner, B. A. (2016). *The Chicago guide to grammar, usage, and punctuation*. University of Chicago Press.
- Garside, R. & Smith, N. (1997). A hybrid grammatical tagger: CLAWS4. In R. Garside, G. Leech, and A. McEnery (Eds.) *Corpus annotation: Linguistic information from computer text corpora*. (pp. 102–121). London: Longman.
- Gilman, E. W. (1994). *Merriam-Webster's dictionary of English usage*. Springfield, MA: Merriam-Webster.
- Gorrell, R. M. (1977). "Usage as rhetoric." *College Composition and Communication*, 28(1), 20–25.
- Graves, H. B. & Graves, R. (1998). Masters, slaves, and infant mortality: Language challenges for technical editing. *Technical Communication Quarterly*, 7(4), 389–414.
- Gray, B. (2015). *Linguistic variation in research articles: When discipline tells only part of the story*. Amsterdam: John Benjamins.

- Gray, B. & Cortes, V. (2011). Perception vs. evidence: An analysis of *this* and *these* in academic prose. *English for Specific Purposes*, 30, 31–43.
- Greenbaum, S. & Whitcut, J. (1988). *Longman guide to English usage*. Harlow, England: Longman.
- Grieve, J. W. (2016). *Regional variation in written American English*. Cambridge University Press.
- Grieve, J. W., Asnaghi, C. & Ruetten, T. (2013). Site-restricted web searches for data collection in regional dialectology. *American Speech* 88(4), 413–440.
- Grieve, J., Biber, D., Friginal, E. & Nekrasova, T. (2011). Variation among blogs: A multi-dimensional analysis. In A. Mehler, S. Sharoff, and M. Santini (Eds.) *Genres on the web: Computational models and empirical studies*. (pp. 303-322). Amsterdam: Springer.
- Hall, J. L. (1917). *English usage: Studies in the history and uses of English words and phrases*. Chicago: Scott Foresman and Co.
- Heylighen, F., & Dewaele, J.-M. (1999). Formality of language: Definition, measurement and behavioral determinants. Internal Report. Center “Leo Apostel,” Free University of Brussels.
- hone in. (n.d.). In *Merriam-Webster.com*. Retrieved July 16, 2018, from [https://www.merriam-webster.com/dictionary/hone in](https://www.merriam-webster.com/dictionary/hone%20in)
- Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of English*. Cambridge University Press.
- Hyland, K. & Feng, K. J. (2017). Is academic writing becoming more informal? *English for Specific Purposes*, 45, 40–51.
- Ilson, R. (1985). Usage problems in British and American English. In S. Greenbaum (Ed.), *The English language today*. (pp. 166–182). Oxford: Pergamon Press.
- Irvine, J. T. (1979). Formality and informality in communicative events. *American Anthropologist*, 81(4), 773–790.
- Jarvis, J. (2012, September 14). Cutting copy desks: Penny-wise or pound-foolish? [Editorial] *Gateway Journalism Review*. Retrieved from <http://gatewayjr.org/2012/09/14/cutting-copy-desks-penny-wise-or-pound-foolish/>
- Jeffries, L. & McIntyre, D. (2010). *Stylistics*. Cambridge University Press.
- Keith, S. (2015). Horseshoes, stylebooks, wheels, poles, and dummies: Objects of editing power in 20th-century newsrooms. *Journalism*, 16(1), 44–60.

- Labov, W. (1966). *The social stratification of English in New York City*. Washington, D. C.: Center for Applied Linguistics.
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press.
- Lakaw, A. (2017). Diachronic shifts in agreement patterns of collective nouns in 19th-century American and British English. *Exploring Recent Diachrony: Corpus Studies of Lexicogrammar and Language Practices in Late Modern English (Studies in Variation, Contacts and Change in English, 18)*. Edited by S. Hoffmann, A. Sand & S. Arndt-Lappe. Helsinki: VARIENG. Retrieved from <http://www.helsinki.fi/varieng/series/volumes/18/lakaw/>
- LaRocque, P. (2012). Don't depend on copy editors for the basics. *Quill*, 100(5), 9.
- Lee, D. Y. W. (2001). Genres, registers, text types, domains, and styles: Clarifying the concepts and navigating a path through the BNC jungle. *Language Learning & Technology* 5(3), 37–72.
- Leech, G., Hundt, M. Mair, C. & Smith, N. (2009). *Change in contemporary English: A grammatical study*. Cambridge University Press.
- Levin, H., Long, S., & Schaffer, C. A. (1981). The formality of the Latinate lexicon in English. *Language and Speech*, 24(2), 161–171.
- Levin, H. & Novak, M. (1991). Frequencies of Latinate and Germanic words in English as determinants of formality. *Discourse Processes*, 14(3), 389–398.
- Lillis, T. (2013). *The sociolinguistics of writing*. Edinburgh University Press.
- Liu, D. & Zhan, H. (2015). The use of the *-free* compound and *free of* and *free from* phrasal constructions: A diachronic and synchronic study. *Journal of English Linguistics*, 43(3), 201–226.
- Lowth, R. (1799). *A short introduction to English grammar*.
- Mackey, A. and Gass, S. (2016). *Second language research: Methodology and design* (2nd ed.). New York: Routledge.
- Mackiewicz, J. (1999). Gauging prescriptivism in writing handbooks. *IEEE Transactions on Professional Communication*, 42(1), 55–58.
- Mackiewicz, J. (2003). Which rules for online writing are worth following?: A study of eight rules in eleven handbooks. *IEEE Transactions on Professional Communication*, 46(2), 129–137.
- Mackiewicz, J. (2010). The co-construction of credibility in online product reviews. *Technical Communication Quarterly*, 19(4), 403–426.

- McIntyre, J. (2017, October 7). Hey, AP Stylebook! Hello? Hello? Anybody there? [Blog post]. Retrieved from <http://www.baltimoresun.com/news/language-blog/bal-hey-ap-stylebook-hello-hello-anybody-there-20171007-story.html>
- McWhorter, J. (1998). *Word on the street: Debunking the myth of "pure" standard English*. Cambridge, Mass: Perseus Publishing.
- Meyers, W. E. (1972). A study of usage items based on an examination of the Brown Corpus. *College Composition and Communication*, 23, 155–169.
- Meyers, W. E. (1995). Linguistics in textbooks: A forty-year comparison. *American Speech*, 70(1), 30–68.
- Miłkowski, M. (2013). On the social nature of linguistic prescriptions. *Psychology of Language and Communication*, 17(2), 175–187.
- Miller, C. R. (1979). A humanistic rationale for technical writing. *College English*, 40(6), 610–617.
- Millward, C. M. (1996). *A biography of the English language* (2nd ed.) Boston, Massachusetts: Thomson Wadsworth.
- Milroy, J. (1992). *Language variation and change*. Oxford: Blackwell.
- Milroy, J. & Milroy, L. (1999). *Authority in language: Investigating standard English* (3rd ed.). New York: Routledge.
- Mittins, W. H., Salu, M., Edminson M., & Coyne, S. (1970). *Attitudes to English usage: An enquiry by the University of Newcastle upon Tyne Institute of Education English Research Group*. London: Oxford University Press.
- Myers, G. (2010). *Discourse of blogs and wikis*. New York: Continuum International Publishing Group.
- Myhill, J. (2004a). Why has Black English not been standardized? A cross-cultural dialogue on prescriptivism. *Language Sciences*, 26, 27–56.
- Myhill, J. (2004b). A parameterized view of the concept of ‘correctness’. *Multilingua*, 23, 389–416.
- Nesfield, J. (1898). *English grammar past and present in three parts*. London: MacMillan & Co.
- O’Connor, P. T. (2009). *Woe is I: The grammarphobe’s guide to better English is plain English*. New York: Riverhead Books.

- Owen, J. (2013). "At the coal-face of standardization": Uncovering the role of copy editors in standardizing the English language (Unpublished master's thesis). Brigham Young University, Provo, Utah.
- Parker, F., & Riley, K. L. (2005). *Grammar for grammarians: Prescriptive, descriptive, generative, contextual*. Parlay Press.
- Payne, G. (1911). *Everyday errors in pronunciation, spelling, and spoken English*. San Francisco: Ricardo J. Orozco.
- Perry, M. J. (2015, April 30). Creative destruction: Newspaper ad revenue continued its precipitous free fall in 2014, and it's likely to continue. [Blog post]. Retrieved from <https://www.aei.org/publication/creative-destruction-newspaper-ad-revenue-continued-its-precipitous-free-fall-in-2014-and-its-likely-to-continue/>
- personal blog. (2013). In *Oxford Dictionary of the Internet (3rd ed.)*. Retrieved December 21, 2018 from <http://www.oxfordreference.com.proxy.lib.iastate.edu/view/10.1093/acref/9780191744150.001.0001/acref-9780191744150-e-4304?rskey=JTcNus&result=1>
- Peters, P. (2004). *The Cambridge guide to English usage*. Cambridge University Press.
- Peters, P. (2006). English usage: Prescription and description. In B. Aarts and A. McMahon (Eds.) *The handbook of English linguistics*. (pp. 759–780). Malden, Mass: Blackwell.
- Peters, P. (2018). The lexicography of English usage. In I. Tiekens Boon van Ostade (Ed.) *English usage guides: History, advice, attitudes*. (pp. 31-49). Oxford University Press.
- Peters, P. & Young, W. (1997). English grammar and the lexicography of usage. *Journal of English Linguistics*, 25(4), 315–331.
- Preston, D. R. (2013). Linguistic insecurity forty years later. *Journal of English Linguistics*, 41(4), 304–331.
- Pullum, G. K. (2004). Ideology, power, and linguistic theory. Talk presented at the annual meeting of the Modern Language Association, December 30, 2004. Retrieved from www.lel.ed.ac.uk/~gpullum/MLA2004.pdf
- Pullum, G. K. (2010). The land of the free and *The elements of style*. *English Today*, 26(2), 34-44.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. New York: Longman.
- quotations in the news. (n.d.) *Associated Press Stylebook Online* (accessed May 22, 2019).
- Rude, C. D. & Eaton, A. (2011). *Technical editing* (5th ed.). Boston: Longman.

- Semuels, A. (2018, January 23). The internet is enabling a new kind of poorly paid hell. *The Atlantic*. Retrieved from <https://www.theatlantic.com/business/archive/2018/01/amazon-mechanical-turk/551192/>
- Siegal, A. M. & Connolly, W. G. (2015). *The New York times manual of style and usage*. Three Rivers Press.
- Simpson, P. (2004). *Stylistics: A resource book for students*. New York: Routledge
- Szmrecsanyi, B. (2019). Register in variationist linguistics. *Register Studies*, 1(1), 76–99.
- Spanogle, H. (2014). Sending editors out the door. *Communication: Journalism Education Today*, 47(4), 47-49.
- Stamper, K. (2014). “What’s a usage dictionary and why should you use one?” Talk presented at the annual meeting of the American Copy Editors Society, Las Vegas, NV.
- Straaijer, R. (2015). The Hyper Usage Guide of English Database User Manual. Retrieved from <https://bridgingtheunbridgeable.files.wordpress.com/2015/-02/huge-manual-2015.pdf>.
- Straaijer, R. (2018). The usage guide: Evolution of a genre. In I. Tieken Boon van Ostade (Ed.) *English usage guides: History, advice, attitudes*. (pp. 11-29). Oxford University Press.
- Strunk, W. & White, E. B. (2009). *The elements of style 50th anniversary edition*. Longman.
- Tagliamonte, S. A. (2016). *Teen talk: The language of adolescents*. Cambridge University Press.
- Thayer, A., Evans, M. B., McBride, A. A., Queen, M., & Spyridakis, J. (2010). I, pronoun: A study of formality in online content. *Journal of Technical Writing and Communication*, 40(4), 447–458.
- they. (n.d.). In *Oxford English dictionary*. Retrieved August 2, 2019, from <https://www-oed-com.proxy.lib.iastate.edu/view/Entry/200700?redirectedFrom=they#eid>
- Tieken-Boon van Ostade I. M. (2013). Studying attitudes to English usage – Investigating prescriptivism in a large research project, *English Today*, 29(4): 3-12.
- Trask, R. L. *Mind the gaffe: A troubleshooter’s guide to English style and usage*. New York: HarperCollins.
- Trudgill, P. (1999). Standard English: What it isn’t. In T. Bex and R. J. Watts (Eds.) *Standard English: The widening debate*. (pp. 117–128). New York: Routledge.

- Truss, L. (2004). *Eats, shoots, and leaves: The zero tolerance approach to punctuation*. New York: Gotham Books.
- Tufte, E. (1983). *Visual display of quantitative information*. Cheshire, CT: Graphics Press.
- Venter, M. (2002). What is correct English? *The English Academy Review*, 19(1), 65–73.
- verbs. (n.d.) *Associated Press Stylebook Online* (accessed May 22, 2019).
- Verdonk, P. (2002). *Stylistics*. Oxford University Press.
- Vox.com. (n.d.). Masthead [Organization chart]. Retrieved from <https://www.vox.com/masthead>
- Walsh, T. & Walsh, N. (1989). Patterns of *who/whom* usage. *American Speech*, 64(3), 284–286.
- Warren, T. (2017). [Review of the book *The Chicago guide to grammar, usage, and punctuation*, by B. A. Garner]. *Technical Communication*, 64(3), 263.
- Weiner, E. (1988). On editing a usage guide. In E.G. Stanley and T. F. Hoad (Eds.) *Words for Robert Burchfield's sixty-fifth birthday*. (pp. 171-183). Cambridge: D. S. Brewer.
- Wolfe, J. (2009). How technical communication textbooks fail engineering students. *Technical Communication Quarterly*, 18(4), 351–375.
- Yáñez-Bouza, N. (2015). *Grammar, rhetoric, and usage in English: Preposition placement 1500–1900*. Cambridge University Press.
- Zwicky, A. (2010, February 25). NomConjObjs [Blog post]. Retrieved from <https://arnoldzwicky.org/2010/02/25/nomconjjobs/>

APPENDIX A: UCREL CLAWS7 TAGSET¹⁸

Tag	Description
APPGE	possessive pronoun, pre-nominal (e.g. my, your, our)
AT	article (e.g. the, no)
AT1	singular article (e.g. a, an, every)
BCL	before-clause marker (e.g. in order (that), in order (to))
CC	coordinating conjunction (e.g. and, or)
CCB	adversative coordinating conjunction (but)
CS	subordinating conjunction (e.g. if, because, unless, so, for)
CSA	as (as conjunction)
CSN	than (as conjunction)
CST	that (as conjunction)
CSW	whether (as conjunction)
DA	after-determiner or post-determiner capable of pronominal function (e.g. such, former, same)
DA1	singular after-determiner (e.g. little, much)
DA2	plural after-determiner (e.g. few, several, many)
DAR	comparative after-determiner (e.g. more, less, fewer)
DAT	superlative after-determiner (e.g. most, least, fewest)
DB	before determiner or pre-determiner capable of pronominal function (all, half)
DB2	plural before-determiner (both)
DD	determiner (capable of pronominal function) (e.g. any, some)
DD1	singular determiner (e.g. this, that, another)
DD2	plural determiner (these, those)
DDQ	wh-determiner (which, what)
DDQGE	wh-determiner, genitive (whose)
DDQV	wh-ever determiner, (whichever, whatever)
EX	existential there
FO	formula
FU	unclassified word
FW	foreign word
GE	germanic genitive marker - (' or's)
IF	for (as preposition)
II	general preposition
IO	of (as preposition)
IW	with, without (as prepositions)
JJ	general adjective
JJR	general comparative adjective (e.g. older, better, stronger)
JJT	general superlative adjective (e.g. oldest, best, strongest)

¹⁸ Reproduced from <http://ucrel.lancs.ac.uk/claws7tags.html>, accessed July 7, 2018.

Tag	Description
JK	catenative adjective (able in be able to, willing in be willing to)
MC	cardinal number, neutral for number (two, three..)
MC1	singular cardinal number (one)
MC2	plural cardinal number (e.g. sixes, sevens)
MCGE	genitive cardinal number, neutral for number (two's, 100's)
MCMC	hyphenated number (40-50, 1770-1827)
MD	ordinal number (e.g. first, second, next, last)
MF	fraction, neutral for number (e.g. quarters, two-thirds)
ND1	singular noun of direction (e.g. north, southeast)
NN	common noun, neutral for number (e.g. sheep, cod, headquarters)
NN1	singular common noun (e.g. book, girl)
NN2	plural common noun (e.g. books, girls)
NNA	following noun of title (e.g. M.A.)
NNB	preceding noun of title (e.g. Mr., Prof.)
NNL1	singular locative noun (e.g. Island, Street)
NNL2	plural locative noun (e.g. Islands, Streets)
NNO	numeral noun, neutral for number (e.g. dozen, hundred)
NNO2	numeral noun, plural (e.g. hundreds, thousands)
NNT1	temporal noun, singular (e.g. day, week, year)
NNT2	temporal noun, plural (e.g. days, weeks, years)
NNU	unit of measurement, neutral for number (e.g. in, cc)
NNU1	singular unit of measurement (e.g. inch, centimetre)
NNU2	plural unit of measurement (e.g. ins., feet)
NP	proper noun, neutral for number (e.g. IBM, Andes)
NP1	singular proper noun (e.g. London, Jane, Frederick)
NP2	plural proper noun (e.g. Browns, Reagans, Koreas)
NPD1	singular weekday noun (e.g. Sunday)
NPD2	plural weekday noun (e.g. Sundays)
NPM1	singular month noun (e.g. October)
NPM2	plural month noun (e.g. Octobers)
PN	indefinite pronoun, neutral for number (none)
PN1	indefinite pronoun, singular (e.g. anyone, everything, nobody, one)
PNQO	objective wh-pronoun (whom)
PNQS	subjective wh-pronoun (who)
PNQV	wh-ever pronoun (whoever)
PNX1	reflexive indefinite pronoun (oneself)
PPGE	nominal possessive personal pronoun (e.g. mine, yours)
PPH1	3rd person sing. neuter personal pronoun (it)
PPHO1	3rd person sing. objective personal pronoun (him, her)
PPHO2	3rd person plural objective personal pronoun (them)
PPHS1	3rd person sing. subjective personal pronoun (he, she)

Tag	Description
PPHS2	3rd person plural subjective personal pronoun (they)
PPIO1	1st person sing. objective personal pronoun (me)
PPIO2	1st person plural objective personal pronoun (us)
PPIS1	1st person sing. subjective personal pronoun (I)
PPIS2	1st person plural subjective personal pronoun (we)
PPX1	singular reflexive personal pronoun (e.g. yourself, itself)
PPX2	plural reflexive personal pronoun (e.g. yourselves, themselves)
PPY	2nd person personal pronoun (you)
RA	adverb, after nominal head (e.g. else, galore)
REX	adverb introducing appositional constructions (namely, e.g.)
RG	degree adverb (very, so, too)
RGQ	wh- degree adverb (how)
RGQV	wh-ever degree adverb (however)
RGR	comparative degree adverb (more, less)
RGT	superlative degree adverb (most, least)
RL	locative adverb (e.g. alongside, forward)
RP	prep. adverb, particle (e.g. about, in)
RPK	prep. adv., catenative (about in be about to)
RR	general adverb
RRQ	wh- general adverb (where, when, why, how)
RRQV	wh-ever general adverb (wherever, whenever)
RRR	comparative general adverb (e.g. better, longer)
RRT	superlative general adverb (e.g. best, longest)
RT	quasi-nominal adverb of time (e.g. now, tomorrow)
TO	infinitive marker (to)
UH	interjection (e.g. oh, yes, um)
VB0	be, base form (finite i.e. imperative, subjunctive)
VBDR	were
VBDZ	was
VBG	being
VBI	be, infinitive (To be or not... It will be ..)
VBM	am
VDN	been
VBR	are
VBZ	is
VD0	do, base form (finite)
VDD	did
VDG	doing
VDI	do, infinitive (I may do... To do...)
VDN	done
VDZ	does

Tag	Description
VH0	have, base form (finite)
VHD	had (past tense)
VHG	having
VHI	have, infinitive
VHN	had (past participle)
VHZ	has
VM	modal auxiliary (can, will, would, etc.)
VMK	modal catenative (ought, used)
VV0	base form of lexical verb (e.g. give, work)
VVD	past tense of lexical verb (e.g. gave, worked)
VVG	-ing participle of lexical verb (e.g. giving, working)
VVGK	-ing participle catenative (going in be going to)
VVI	infinitive (e.g. to give... It will work...)
VVN	past participle of lexical verb (e.g. given, worked)
VVNK	past participle catenative (e.g. bound in be bound to)
VVZ	-s form of lexical verb (e.g. gives, works)
XX	not, n't
ZZ1	singular letter of the alphabet (e.g. A,b)
ZZ2	plural letter of the alphabet (e.g. A's, b's)

APPENDIX B: SURVEY INSTRUMENT

Demographic Information

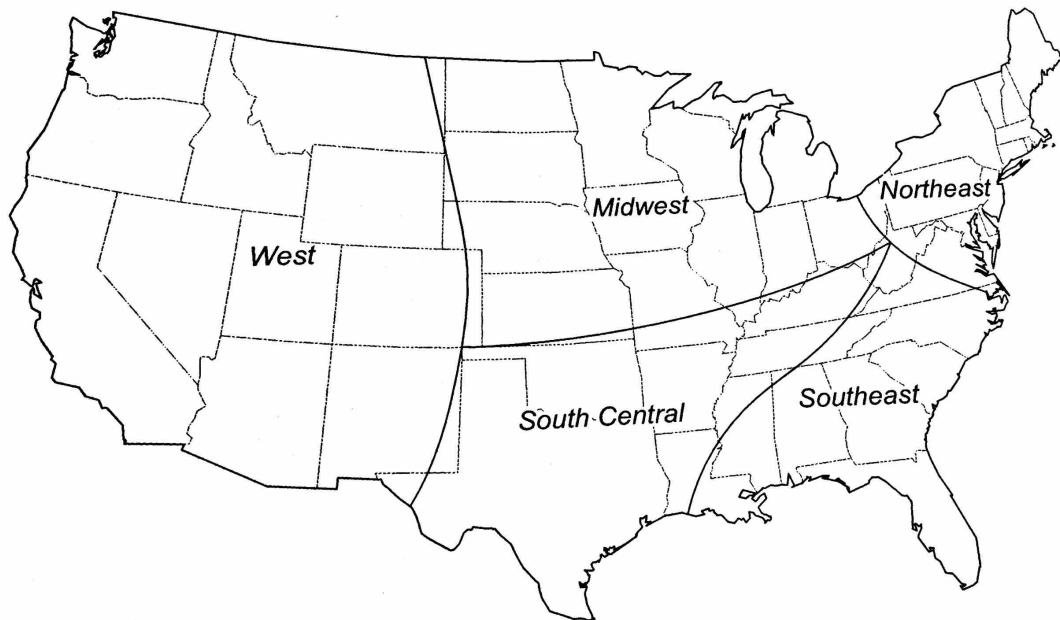
Please answer the following questions. The information in this section allows me to learn more about factors that may influence the kind of the language you use when you write posts for your blog (bloggers)/news articles (journalists)

Age

Gender

Occupation (bloggers)/Title (journalists)

Please refer to this map to answer the three questions that follow:



Source: Grieve, J. W. (2016). *Regional variation in written American English*. Cambridge University Press. (Used with permission)

In which region did you grow up?

- West
- Midwest
- South Central
- Southeast
- Northeast

In which region have you lived the longest?

- West
- Midwest
- South Central
- Southeast
- Northeast

In which region do you currently live?

- West
- Midwest
- South Central
- Southeast
- Northeast

Experience with English Reference Materials

Please provide brief answers to the following questions. Keep in mind there are no right or wrong answers.

Are you familiar with any English-language reference materials (for example, grammar books, dictionaries, style manuals, usage guides, language-related websites, etc.)?

- Yes
- No

Please list the English-language reference materials you are familiar with. If you are familiar with a lot of English-language reference materials, please list the first 5 that come to mind.

Of those reference materials you mentioned above, are there any that you regularly read or consult?

- Yes
- No

What English-language reference materials do you regularly read or consult?

What are the primary reasons you consult this guide/these reference materials instead of others?

Composing Process

In the following questions, please select the answer that best represents your experiences, feelings, and opinions. There are no right or wrong answers!

Bloggers	Journalists
<p>Which one of the following statements most accurately describes your writing process when composing a blog post? Even if none of the options reflect your writing process with 100% accuracy, please choose the one that is closest.</p> <ol style="list-style-type: none"> I write a post then publish it without reading over what I've written. I write a post then quickly glance over what I've written so I can edit any major mistakes before I publish it. I write a post then carefully edit my writing before I publish it. I carefully edit a post as I write it, then I publish it. I write a post then ask someone else to edit it for me before I publish it. 	<p>Which one of the following statements most accurately describes your writing process when composing a news article? Even if none of the options reflect your writing process with 100% accuracy, please choose the one that is closest.</p> <ol style="list-style-type: none"> I write an article then submit it without reading over what I've written. I write an article then quickly glance over what I've written so I can edit any major mistakes before I submit it. I write an article then carefully edit my writing before I submit it. I carefully edit an article as I write it, then I submit it. I write an article then someone else edits it for me before it is submitted.
<p>How much effort do you put into checking your blog posts to make sure they are grammatically correct?</p> <ol style="list-style-type: none"> A great deal Some Only a little None at all 	<p>How much effort do you put into checking your news articles to make sure they are grammatically correct?</p> <ol style="list-style-type: none"> A great deal Some Only a little None at all
<p>How important is it to you that your blog posts be completely free of any grammatical errors?</p> <ol style="list-style-type: none"> Very important Somewhat important Slightly important Not at all important 	<p>How important is it to you that your articles be completely free of any grammatical errors?</p> <ol style="list-style-type: none"> Very important Somewhat important Slightly important Not at all important
<p>To what extent do you agree with the following statement: The writing found in personal blogs online should have the same level of grammatical correctness as news writing.</p>	<p>To what extent do you agree with the following statement: News writing and the writing found in personal blogs online should have the same level of grammatical correctness.</p>

<ul style="list-style-type: none"> a. Completely agree b. Generally agree c. Generally disagree d. Completely disagree 	<ul style="list-style-type: none"> a. Completely agree b. Generally agree c. Generally disagree d. Completely disagree
<p>When writing posts for your blog, how often does another person (e.g., friend, family member—someone who is not a professional editor) edit your work before it's published?</p> <ul style="list-style-type: none"> a. Always b. Usually c. Sometimes d. Rarely e. Never 	<p>When writing news articles, how often does another person (e.g., friend, family member—someone who is not a professional editor) edit your work before it's published?</p> <ul style="list-style-type: none"> a. Always b. Usually c. Sometimes d. Rarely e. Never
<p>How often does a professional editor edit your blog posts before they're published?</p> <ul style="list-style-type: none"> a. Always b. Usually c. Sometimes d. Rarely e. Never 	<p>How often does a professional editor edit your articles before they're published?</p> <ul style="list-style-type: none"> a. Always b. Usually c. Sometimes d. Rarely e. Never
<p>How often do you use an automated spell checker to help you edit your blog posts before you publish them?</p> <ul style="list-style-type: none"> a. Always b. Usually c. Sometimes d. Rarely e. Never 	<p>How often do you use an automated spell checker to help you edit your articles before you publish them?</p> <ul style="list-style-type: none"> a. Always b. Usually c. Sometimes d. Rarely e. Never
<p>How often do you use an automated grammar checker to help you edit your blog posts before you publish them?</p> <ul style="list-style-type: none"> f. Always g. Usually h. Sometimes i. Rarely j. Never 	<p>How often do you use an automated grammar checker to help you edit your articles before you publish them?</p> <ul style="list-style-type: none"> f. Always g. Usually h. Sometimes i. Rarely j. Never
<p>When you're writing a blog post and you're not sure how to use a particular word, how often do you look it up in some kind of</p>	<p>When you're writing an article and you're not sure how to use a particular word, how often do you look it up in some kind of</p>

reference material (e.g., dictionary, usage guide, style manual, etc.)? a. Always b. Usually c. Sometimes d. Rarely e. Never	reference material (e.g., dictionary, usage guide, style manual, etc.)? a. Always b. Usually c. Sometimes d. Rarely e. Never
In general, how would you classify the writing in your blog posts? a. Very formal b. Somewhat formal c. Somewhat informal d. Very informal	In general, how would you classify the writing in your articles? a. Very formal b. Somewhat formal c. Somewhat informal d. Very informal

Attitudes to English Usage

The following two questions ask about your feelings toward the English language. Again, there are no right or wrong answers!

What do you think about the state of the English language today?

Do you have any pet peeves when it comes to language use? In other words, are there things that you hear people say or things that you see people write that bother you? If so, what are some of the most important ones to you?

Acceptability Judgments

This section contains 12 short sentences, 19 each one followed by a series of questions. Please answer the questions according to what you consider acceptable in your own language use. Would you say or write these sentences? If so, in which contexts? If not, why not? Note that this is NOT a test! There are no “correct” answers. I am just interested in

¹⁹ Sentences are the same as or adapted from those contained in the HUGE database.

what you think about these sentences. Additionally, please go through the questions as quickly as possible, as your initial opinion is what I am hoping to get.

1. When my sister's baby is born, I will be an uncle.

Is this sentence grammatically acceptable or unacceptable? If acceptable, in which context(s)?

- Unacceptable
- Acceptable in formal writing
- Acceptable in formal speaking
- Acceptable in formal online communication
- Acceptable in informal writing
- Acceptable in informal speaking
- Acceptable in informal online communication

How certain do you feel about its acceptability/unacceptability?

- Very certain
- Somewhat certain
- Somewhat uncertain
- Very uncertain

Why did you choose this answer?

- You used a rule
- It sounded right/wrong

Any comments?

2. The tools were just laying there.

Is this sentence grammatically acceptable or unacceptable? If acceptable, in which context(s)?

- Unacceptable
- Acceptable in formal writing
- Acceptable in formal speaking
- Acceptable in formal online communication
- Acceptable in informal writing
- Acceptable in informal speaking
- Acceptable in informal online communication

How certain do you feel about its acceptability/unacceptability?

- Very certain
- Somewhat certain
- Somewhat uncertain
- Very uncertain

Why did you choose this answer?

- You used a rule
- It sounded right/wrong

Any comments?

3. Who did she ask?

Is this sentence grammatically acceptable or unacceptable? If acceptable, in which context(s)?

- Unacceptable
- Acceptable in formal writing
- Acceptable in formal speaking
- Acceptable in formal online communication
- Acceptable in informal writing
- Acceptable in informal speaking
- Acceptable in informal online communication

How certain do you feel about its acceptability/unacceptability?

- Very certain
- Somewhat certain
- Somewhat uncertain
- Very uncertain

Why did you choose this answer?

- You used a rule
- It sounded right/wrong

Any comments?

4. Cats are very different than dogs.

Is this sentence grammatically acceptable or unacceptable? If acceptable, in which context(s)?

- Unacceptable
- Acceptable in formal writing
- Acceptable in formal speaking
- Acceptable in formal online communication
- Acceptable in informal writing
- Acceptable in informal speaking
- Acceptable in informal online communication

How certain do you feel about its acceptability/unacceptability?

- Very certain
- Somewhat certain
- Somewhat uncertain
- Very uncertain

Why did you choose this answer?

- You used a rule
- It sounded right/wrong

Any comments?

5. She refused to even think about it.

Is this sentence grammatically acceptable or unacceptable? If acceptable, in which context(s)?

- Unacceptable
- Acceptable in formal writing
- Acceptable in formal speaking
- Acceptable in formal online communication
- Acceptable in informal writing
- Acceptable in informal speaking
- Acceptable in informal online communication

How certain do you feel about its acceptability/unacceptability?

- Very certain
- Somewhat certain
- Somewhat uncertain
- Very uncertain

Why did you choose this answer?

- You used a rule
- It sounded right/wrong

Any comments?

6. He only had one chapter to finish.

Is this sentence grammatically acceptable or unacceptable? If acceptable, in which context(s)?

- Unacceptable
- Acceptable in formal writing
- Acceptable in formal speaking

- Acceptable in formal online communication
- Acceptable in informal writing
- Acceptable in informal speaking
- Acceptable in informal online communication

How certain do you feel about its acceptability/unacceptability?

- Very certain
- Somewhat certain
- Somewhat uncertain
- Very uncertain

Why did you choose this answer?

- You used a rule
- It sounded right/wrong

Any comments?

7. **She told Charles and I the whole story.**

Is this sentence grammatically acceptable or unacceptable? If acceptable, in which context(s)?

- Unacceptable
- Acceptable in formal writing
- Acceptable in formal speaking
- Acceptable in formal online communication
- Acceptable in informal writing
- Acceptable in informal speaking
- Acceptable in informal online communication

How certain do you feel about its acceptability/unacceptability?

- Very certain
- Somewhat certain
- Somewhat uncertain
- Very uncertain

Why did you choose this answer?

- You used a rule
- It sounded right/wrong

Any comments?

8. **Everyone has their own style.**

Is this sentence grammatically acceptable or unacceptable? If acceptable, in which context(s)?

- Unacceptable
- Acceptable in formal writing
- Acceptable in formal speaking
- Acceptable in formal online communication
- Acceptable in informal writing
- Acceptable in informal speaking
- Acceptable in informal online communication

How certain do you feel about its acceptability/unacceptability?

- Very certain
- Somewhat certain
- Somewhat uncertain
- Very uncertain

Why did you choose this answer?

- You used a rule
- It sounded right/wrong

Any comments?

9. **There were less accidents this year than last year.**

Is this sentence grammatically acceptable or unacceptable? If acceptable, in which context(s)?

- Unacceptable
- Acceptable in formal writing
- Acceptable in formal speaking
- Acceptable in formal online communication
- Acceptable in informal writing
- Acceptable in informal speaking
- Acceptable in informal online communication

How certain do you feel about its acceptability/unacceptability?

- Very certain
- Somewhat certain
- Somewhat uncertain
- Very uncertain

Why did you choose this answer?

- You used a rule

- It sounded right/wrong

Any comments?

10. None were left on the table.

Is this sentence grammatically acceptable or unacceptable? If acceptable, in which context(s)?

- Unacceptable
- Acceptable in formal writing
- Acceptable in formal speaking
- Acceptable in formal online communication
- Acceptable in informal writing
- Acceptable in informal speaking
- Acceptable in informal online communication

How certain do you feel about its acceptability/unacceptability?

- Very certain
- Somewhat certain
- Somewhat uncertain
- Very uncertain

Why did you choose this answer?

- You used a rule
- It sounded right/wrong

Any comments?

11. I've seen two opera's this summer.

Is this sentence grammatically acceptable or unacceptable? If acceptable, in which context(s)?

- Unacceptable
- Acceptable in formal writing
- Acceptable in formal speaking
- Acceptable in formal online communication
- Acceptable in informal writing
- Acceptable in informal speaking
- Acceptable in informal online communication

How certain do you feel about its acceptability/unacceptability?

- Very certain

- Somewhat certain
- Somewhat uncertain
- Very uncertain

Why did you choose this answer?

- You used a rule
- It sounded right/wrong

Any comments?

12. That joke was so funny that I literally died laughing.

Is this sentence grammatically acceptable or unacceptable? If acceptable, in which context(s)?

- Unacceptable
- Acceptable in formal writing
- Acceptable in formal speaking
- Acceptable in formal online communication
- Acceptable in informal writing
- Acceptable in informal speaking
- Acceptable in informal online communication

How certain do you feel about its acceptability/unacceptability?

- Very certain
- Somewhat certain
- Somewhat uncertain
- Very uncertain

Why did you choose this answer?

- You used a rule
- It sounded right/wrong

Any comments?

Thank you for participating in this survey! Your responses have been recorded. If you have any questions about the project, please feel free to contact Jordan Smith at tjordans@iastate.edu.

APPENDIX C: PRESCRIPTIVISM RATINGS FOR EACH USAGE PROBLEM

	DIFFERENT TO/THAN/FROM	I FOR ME	LAY/LIE	LESS/FEWER	NONE IN PLURAL CONTEXT	SINGULAR THEY	SPLIT INFINITIVE	WHO/WHOM
Brians (2013)	2	4	3.5	4	1	2	2.5	2
Butterfield (2015)	1.5	3	3	3	2	1.5	3	2
Garner (2016)	3	4	4	3	1	2.5	2.5	2.5
Fogarty (2008)	4	4	4	3.5	--	2.5	1	4
MWDEU (1994)	1	2.5	2	3	1	1	2	2
Trask (2006)	3	3.5	4	3	1	3	1	1.5
AHGCUS (2005)	3	2.5	2.5	3	2	3	1.5	3
Peters (2004)	1.5	3	3	2.5	1	1	2	2
Strunk and White (2009)	4	4	3.5	4	2	4	3	4
Batko (2004)	3	4	4	4	2	3	--	4
O'Connor (2009)	3	4	4	3.5	1.5	4	1	2
Avg	2.64	3.50	3.41	3.32	1.45	2.50	1.95	2.64
StDev	1.00	0.63	0.70	0.51	0.50	1.05	0.80	0.95
Min	1	2.5	2	2.5	1	1	1	1.5
Max	4	4	4	4	2	4	3	4

APPENDIX D: WALSH AND WALSH'S (1989) LIST OF GRAMMATICAL PATTERNS
FOR *WHO/WHOM*

1. Object of a preposition in main clause with the preposition stranded
2. Subject of a relative clause
3. Subject of a bare infinitive clause embedded in a relative clause
4. Subject of a main clause
5. Direct Object of a relative clause
6. Subject of a tensed clause embedded in a relative clause
7. Direct object of a main clause
8. Object of a preposition in a relative clause with the preposition fronted
9. Subject of a tensed clause embedded in a main clause
10. Object of a preposition in a main clause with the preposition fronted
11. Object of a preposition in a relative clause with the preposition stranded
12. Subject of a bare infinitive clause embedded in the main clause

APPENDIX E: REGULAR EXPRESSIONS USED TO EXTRACT DATA FOR SINGULAR THEY

	Subject	Object	Possessive	Reflexive
Masculine	\bhe_\w+[@%]?	\bhim_\w+[@%]?	\bhis_\w+[@%]?	\bhimsel_\w+[@%]?
Feminine	\bshe_\w+[@%]?	\bher_\w+[@%]?	\bher_\w+[@%]? \bhers_\w+[@%]?	\bhersel_\w+[@%]?
Indefinite	\bone_PN1[@%]?	\bone_PN1[@%]?	\bone_\w+[@%]? 's_\w+[@%]?	\bonesel_\w+[@%]?
Combined	(?:\bs/he_\w+[@%]? \bhe_\w+[@%]? or_\w+[@%]? she_\w+[@%]? \bshe_\w+[@%]? or_\w+[@%]? he_\w+[@%]? \bhe/she_\w+[@%]? \bshe/he_\w+[@%]?)	(?:\bhim_\w+[@%]? or_\w+[@%]? her_\w+[@%]? \bher_\w+[@%]? or_\w+[@%]? him_\w+[@%]? \bhim/her_\w+[@%]? \bher/him_\w+[@%]?)	(?:\bhis_\w+[@%]? or_\w+[@%]? her_\w+[@%]? \bhis_\w+[@%]? or_\w+[@%]? hers_\w+[@%]? \bher_\w+[@%]? or_\w+[@%]? his_\w+[@%]? \bhers_\w+[@%]? or_\w+[@%]? his_\w+[@%]? \bhis/her_\w+[@%]? \bhis/hers_\w+[@%]? \bher/his_\w+[@%]? \bher/s/his_\w+[@%]?)	(?:\bhis_\w+[@%]? or_\w+[@%]? hersel_\w+[@%]? \bher_\w+[@%]? or_\w+[@%]? himself_\w+[@%]? \bhim_\w+[@%]? or_\w+[@%]? hersel_\w+[@%]? \bhimsel_\w+[@%]? or_\w+[@%]? hersel_\w+[@%]? \bhersel_\w+[@%]? or_\w+[@%]? himself_\w+[@%]? \bhimsel/hersel_\w+[@%]? \bhersel/himself_\w+[@%]?)
Plural	\bthey_\w+[@%]?	\bthem_\w+[@%]?	(?:\btheir_\w+[@%]? \btheirs_\w+[@%]?)	(?:\bthemselves_\w+[@%]? \bthemsel_\w+[@%]?)